Numero 299
Luglio 1993
Anno XXVI
Volume LI

L'economia della vita e della morte

Le statistiche sulla mortalità possono essere gli indicatori più adeguati della reale situazione economica di un paese, in quanto richiamano l'attenzione anche su fondamentali aspetti dell'organizzazione sociale

di Amartya Sen

L'economia non si occupa soltanto di reddito e di ricchezza, ma anche del modo in cui queste risorse possono essere impiegate per conseguire obiettivi significativi, compresa la possibilità di godere di una vita lunga e degna di essere vissuta. Ma se, come spesso avviene, si giudica il successo economico di una nazione solo sulla base del reddito e degli altri tradizionali indicatori di solidità finanziaria, si perde di vista l'obiettivo del benessere. I criteri convenzionali con cui si misura il successo economico potrebbero essere migliorati stimando la capacità che una nazione ha di aumentare la durata della vita e di migliorarne la qualità.

Se è vero che il mondo nel suo complesso ha raggiunto una prosperità senza precedenti, è anche vero che molte zone sono ancora tormentate dalle carestie e da una cronica sottoalimentazione. Non solo nel Terzo Mondo, ma anche in paesi industrializzati, è diffuso il flagello di malattie e morti per denutrizione che potrebbero essere scongiurate. Per risolvere questi problemi è necessario assumere punti di vista nuovi in campo economico. Affiancando agli indicatori tradizionali statistiche più direttamente correlate al benessere, si potrebbero valutare pro e contro di metodi economici alternativi. Per esempio, un paese che abbia un prodotto nazionale lordo (PNL) pro capite molto superiore a quello di un altro, potrebbe avere allo stesso tempo una speranza di vita inferiore a causa di una minore possibilità di accesso al servizio sanitario e all'istruzione di base. I dati sulla mortalità possono servire per giudicare le iniziative politiche e per identificare aspetti fondamentali della povertà economica nei diversi contesti.

Per illustrare l'importanza e il merito delle statistiche sulla mortalità, si possono prendere in considerazione diversi problemi diffusi nel mondo: le carestie, per esempio, che spesso producono i loro effetti devastanti anche quando il cibo sarebbe facilmente disponibile; la ridotta speranza di vita, riscontrabile anche in paesi con un elevato prodotto nazionale lordo; gli indici di mortalità femminili superiori a quelli maschili in regioni asiatiche e africane; e i bassissimi indici di sopravvivenza degli afro-americani, non solo in confronto ai bianchi degli Stati Uniti, ma anche a popolazioni di paesi estremamente poveri.

Spesso per cercare spiegazioni economiche delle carestie si misurano la produzione e la disponibilità di alimenti. E le scelte politiche si basano di frequente su statistiche aggregate a livello nazionale della quantità di prodotti alimentari disponibile per persona, un indicatore portato in auge da Thomas R. Malthus ai primi del secolo scorso. Eppure, contrariamente a quanto si crede, possono verificarsi carestie anche quando l'indicatore globale è alto. Affidarsi a cifre così semplicistiche può generare un'illusoria sicurezza che impedisce ai Governi di prendere misure contro le carestie.

Studentesse in una strada dello Stato indiano del Kerala. Il Kerala, pur avendo un prodotto lordo pro capite tra i più bassi del paese, ha un alto tasso di alfabetizzazione per entrambi sessi. Grazie all'attenzione del Governo per l'istruzione pubblica, per la sanità e per una migliore condizione delle donne, la popolazione del Kerala è generalmente alfabetizzata e longeva nonostante l'estrema povertà. Questo fatto dimostra che indicatori di benessere economico come il PNL possono essere insufficienti.

Per capire meglio le carestie, bisogna tanto esaminare i canali di acquisto e di distribuzione dei generi alimentari, quanto considerare di quali diritti godano i diversi segmenti della società. La fame si diffonde quando a una cospicua frazione di popolazione mancano i mezzi di sostentamento, e questo può avvenire a causa della disoccupazione, della caduta del potere d'acquisto dei salari o di una modifica della ragione di scambio tra beni e servizi venduti e generi alimentari acquistati. Le iniziative politiche volte a prevenire carestie e sottoalimentazione dovrebbero fondarsi su informazioni relative a questi fattori e agli altri processi economici che influiscono sulla capacità che un particolare gruppo ha di procurarsi prodotti alimentari.

La carestia del 1974 nel Bangladesh sta a dimostrare la necessità di valutare con maggior cura le cause di questo tipo di calamità. Quell'anno, nel Bangladesh, la quantità di cibo pro capite era alta: superiore, anzi, a quella di tutti gli altri anni compresi tra il 1971 e il 1976. Ma le inondazioni susseguitesi dalla fine di giugno sino ad agosto interferirono con i lavori di trapianto del riso (cioè con il trasferimento delle piantine dai luoghi di semina, sparsi qua e là, a file ben ordinate in campi coperti d'acqua) e con altre attività agricole nel distretto settentrionale. Quello scompiglio, a sua volta, produsse un aumento della disoccupazione tra i contadini, che per lo più vivono alla giornata. Privi di salario, questi lavoratori non potevano più provvedere al proprio sostentamento e caddero vittime della fame.

Il panico aggravò la situazione. Nelle principali piantagioni di riso, che erano state solo parzialmente danneggiate dalle inondazioni, il raccolto era atteso solo per dicembre, ma la previsione di una produzione insufficiente portò subito all'accumulo di scorte per ragioni precauzionali e speculative. In tutto il paese salirono bruscamente i prezzi del riso e degli altri cereali, con il conseguente crollo del potere d'acquisto dei salari degli abitanti del Bangladesh. In ottobre, quando i prezzi dei prodotti alimentari toccarono il culmine, si registrò anche il più alto tributo in vite umane.

A questo punto, il Governo si mosse tardivamente con un programma di aiuti su larga scala. Il ritardo era dovuto a diverse ragioni, tra cui la sospensione degli aiuti alimentari da parte degli Stati Uniti a causa di un contenzioso sull'esportazione di juta dal Bangladesh a Cuba. Uno degli ostacoli principali, però, fu l'illusorio senso di sicurezza prodotto da cifre rassicuranti sulla disponibilità di generi alimentari. Una volta messi in moto i soccorsi, il mercato iniziò a riallinearsi a una stima più realistica della scarsità prevista per l'inverno: la diminuzione dei raccolti era molto più modesta di quanto valutato in precedenza. Da novembre in poi, i prezzi cominciarono a diminuire, e molti centri di soccorso vennero chiusi alla fine del mese. La carestia era in gran parte superata prima ancora che venisse avviato il raccolto delle messi parzialmente danneggiate.

Come si diceva, quell'anno in Bangladesh la quantità di prodotti alimentari

pro capite era elevata (dato l'ottimo raccolto del dicembre 1973). Questa carestia dimostra quali disastrose conseguenze si possano avere se ci si affida unicamente al valore numerico dell'offerta di generi alimentari. Questi infatti non sono mai equamente divisi tra tutta la popolazione; inoltre, le scorte private e commerciali di prodotti sono offerte o ritirate dal mercato in risposta a incentivi monetari e alle aspettative sulle variazioni dei prezzi.

Spesso una carestia ha luogo senza che le statistiche abbiano previamente registrato una riduzione nell'offerta di generi alimentari. La carestia in Bengala nel 1943, per esempio, ebbe inizio con la diminuzione del potere d'acquisto dei salari dei contadini. Analogamente, nel 1973 una carestia nella provincia etiope del Wollo fu provocata da una grave siccità che impoverì la popolazione locale, ma non ridusse in modo sostanziale la produzione alimentare complessiva della nazione. I prezzi erano spesso più bassi nel Wollo che altrove nel paese, data la ridotta capacità d'acquisto della popolazione di quella provincia; parte dei prodotti, addirittura, venne trasferita dalla regione colpita da carestia ad altre zone più ricche. (Questa tragica catena di eventi si verificò anche negli anni quaranta del secolo scorso, quando ingenti derrate alimentari vennero trasferite dall'Irlanda affamata alla ricca Inghilterra.)

Ci sono diversi modi per prevenire le carestie. In Africa e in Asia servirebbe ovviamente una crescita nella produzione di generi alimentari, non solo perché ridurrebbe il costo dei prodotti, ma anche perché aumenterebbe le possibilità economiche di popolazioni in larga misura impiegate in quel genere di produzione. Per rilanciare la produzione sarebbe necessario incentivare gli investimenti in aziende agricole efficienti; inoltre, sarebbero necessarie iniziative politiche volte a potenziare l'irrigazione e a incoraggiare l'innovazione tecnologica (molto trascurata in Africa).

L'aumento della produzione alimentare, però, non è l'unica risposta. La variabilità delle condizioni meteorologiche, anzi, fa sì che un'eccessiva concentrazione delle risorse di una nazione nello sforzo di produrre più cibo possa rendere più vulnerabile la popolazione a siccità e inondazioni. Nell'Africa subsahariana, in particolare, è quanto mai necessario diversificare la produzione, prevedendo la graduale espansione delle attività manifatturiere. Se la gente ha i mezzi economici, in caso di necessità i prodotti alimentari possono essere acquistati all'estero.

La carestia del 1974 in Bangladesh si verificò nonostante una disponibilità alimentare superiore a quella di ogni altro anno tra il 1971 e il 1976. (L'anno di riferimento per il confronto delle disponibilità annue è il 1967.)

Nel Maharashtra, in India, si limitarono gli effetti di una siccità varando un programma di lavori pubblici per distribuire reddito. Si ebbe un calo generalizzato nel consumo di cereali e la penuria fu ripartita fra tutti.

Nel 1983-1984 la produzione alimentare in Botswana, Zimbabwe e Capo Verde fu inferiore all'attesa, ma gli effetti della carestia furono contenuti grazie a programmi pubblici. Sudan ed Etiopia, pur con un calo produttivo minore, soffrirono maggiormente.

Abbiano o no successo i programmi di aumento e diversificazione della produzione, in molti paesi africani e asiatici milioni di persone continueranno a trovarsi sotto la minaccia di inondazioni, siccità e altre catastrofi naturali. In queste situazioni, la carestia può essere evitata aumentando il potere d'acquisto dei gruppi più colpiti, quelli con minore capacità di sostentamento. Il varo di programmi di lavori pubblici può essere un mezzo rapido per distribuire reddito. I nuovi lavoratori assunti per queste attività entrerebbero nel mercato e verrebbe assicurata loro una quota dell'offerta totale di generi alimentari. Naturalmente, la creazione di nuovi posti di lavoro provoca un aumento dei prezzi: invece di lasciare che gli indigenti muoiano di fame, questa iniziativa fa crescere la domanda totale di generi alimentari. L'aumento può essere realmente benefico, in quanto provoca una riduzione dei consumi in altri gruppi relativamente più agiati. Questo processo distribuisce su un maggior numero di persone gli effetti della carestia, in modo da attutirli.

Questi programmi di lavori pubblici non dovrebbero in genere pesare eccessivamente sulle finanze di una nazione povera. Le vittime delle carestie, anche se numerose in termini assoluti, costituiscono normalmente una piccola percentuale della società: di solito meno del 5-10 per cento della popolazione. Dato, poi, che coloro che muoiono di fame sono anche tra le persone più povere, la loro quota di reddito o di consumi alimentari è spesso compresa tra il 2 e il 4 per cento. Ne consegue che non sono necessarie entrate fiscali inesigibili per ricostituire il loro reddito.

Ci sono diversi esempi di successo nella prevenzione della fame tramite iniziative di lavori pubblici. Nello Stato indiano del Maharashtra, tra il 1972 e il 1973, una serie di gravi episodi di siccità portò a una diffusa disoccupazione nel settore agricolo e al dimezzamento della produzione di generi alimentari. Alcuni programmi di lavori pubblici, come la costruzione di strade e di pozzi, salvarono dalla fame i lavoratori colpiti, i quali poterono poi competere con gli altri nell'acquisto dei prodotti. Sebbene la quantità media pro capite di alimenti disponibili fosse a quel tempo molto inferiore a quella dei paesi del Sahel (Alto Volta, ora Burkina Faso, Mauritania, Mali, Niger, Ciad e Senegal), nel Maharashtra la fame non colpì duramente. Nel Sahel, invece, dove non si ebbe un'equa ripartizione dei prodotti, le morti per inedia furono numerose.

Negli ultimi anni, l'India è riuscita a evitare le carestie soprattutto applicando questo tipo di metodo. L'ultima carestia grave si ebbe nel 1943, quattro anni prima dell'indipendenza. Drammatici episodi di scarsità alimentare si verificarono a causa di calamità naturali negli anni 1967, 1973, 1979 e 1987, ma si riuscì a limitarne gli effetti ristabilendo il potere d'acquisto dei segmenti di popolazione più minacciati.

Prevenire le carestie distribuendo reddito monetario è molto diverso dalla diffusa pratica di raccogliere la gente in campi di soccorso per cercare di alimentarla. Questo metodo, usato spesso in Africa, è di solito più lento e può costituire un impegno organizzativo insostenibile per le autorità pubbliche. Inoltre, ammassare la gente in campi lontani dai luoghi di residenza può gettare nel disordine le normali operazioni economiche, quali la coltivazione dei campi e il governo degli animali, minando di conseguenza la produzione futura. Il raggruppamento può anche sconvolgere la vita familiare. Infine, da non dimenticare, i campi diventano spesso terreni di coltura per malattie infettive.

All'opposto, pagare salari per lavori pubblici non minaccia il benessere economico e sociale degli assistiti. Consolida i meccanismi produttivi e di mercato e incrementa l'efficienza di commercianti e trasportatori. Di fatto questo metodo può rafforzare, invece che indebolire, le infrastrutture economiche.

Una vittima somala della carestia, in attesa di cibo. Le guerre locali, e il venir meno della legalità e dell'ordine, hanno completamente distrutto l'economia somala, provocando l'impoverimento di molte persone. Le precedenti dittature militari poco hanno fatto per prevenire le carestie: la soppressione dei partiti di opposizione e l'eliminazione della libera stampa consentivano infatti a questi regimi di agire in modo irresponsabile.

Inevitabilmente, le iniziative pubbliche di sostegno sono strettamente correlate alle scelte politiche. Per quanto faccia affidamento sul mercato, il metodo dei lavori pubblici non è una politica di mercato libero, in quanto richiede che il Governo intervenga con un'offerta di lavoro. Può essere utile anche la proprietà pubblica di almeno un minimo di scorte alimentari, in modo che il Governo abbia un'arma credibile per contrastare i tentativi dei commercianti di manipolare il mercato: in caso di artificiale rarefazione della merce per far rialzare i prezzi, il Governo può rispondere immettendo derrate sul mercato per provocare un crollo dei prezzi e dei profitti.

La carestia è del tutto evitabile se il Governo è incentivato ad agire tempestivamente. È indicativo il fatto che nessun paese democratico con una stampa relativamente libera abbia mai dovuto affrontare una grave carestia (anche se certi paesi hanno agito più efficacemente di altri sulla via della prevenzione). Questo vale sia per le democrazie ricche sia per quelle povere. Una carestia può portare alla morte di milioni di persone, ma raramente raggiunge i governanti. Se i leader politici devono riconquistarsi l'elezione, e se la stampa è libera di parlare della situazione e di criticare le scelte politiche, allora i responsabili governativi sono incentivati ad agire in termini di prevenzione. In India, per esempio, le carestie sono cessate con l'indipendenza: un sistema democratico pluripartitico e una stampa relativamente indipendente hanno obbligato il Governo ad agire. Al contrario, per quanto molto più avanti dell'India nell'espansione economica e nell'assistenza sanitaria, la Cina postrivoluzionaria non è stata capace di sconfiggere definitivamente le carestie. Una si ebbe tra il 1958 e il 1961, dopo il fallimento del programma agricolo legato al «grande balzo avanti». L'assenza di opposizione politica e di libera stampa permise che si proseguisse per altri tre anni in quella disastrosa iniziativa politica. Ne conseguì un tributo di vite compreso tra 23 e 30 milioni di persone.

Molti paesi dell'Africa subsahariana, tra cui la Somalia, l'Etiopia e il Sudan, hanno pagato un prezzo altissimo per i rispettivi regimi militari. I conflitti e le

guerre portano alla carestia non solo per la loro distruttività economica, ma anche perché favoriscono l'affermarsi di dittature e la diffusione della censura. Paesi subsahariani relativamente democratici, come Botswana e Zimbabwe, sono generalmente riusciti a prevenire le carestie. Naturalmente, anche un paese non democratico può riuscire fortunosamente a evitare una carestia: può capitare che non si verifichino crisi o che un dittatore illuminato attui politiche di sostegno efficaci. Ma la democrazia è la migliore garanzia di azione tempestiva.

I dati sulla mortalità da carestia fanno soffermare l'attenzione sui fallimenti di certe strutture economiche e politiche. Indici di mortalità cronicamente elevati rivelano fallimenti meno drammatici, ma più persistenti. Le politiche economiche associate a una bassa mortalità infantile e a una crescente speranza di vita sono di vario tipo. Diversi paesi in cui si è avuta una drastica riduzione della mortalità infantile negli anni tra il 1960 e il 1985 hanno registrato una crescita economica di rapidità senza precedenti. È il caso di Hong-Kong, Singapore e della Corea del Sud, paesi attualmente ricchi in termini di PNL. Ma nella classifica del successo si trovano anche nazioni tuttora povere: Cina, Giamaica e Costa Rica, tra le altre.

Il fatto che un paese povero possa ottenere, nell'assistenza sanitaria e nella speranza di vita, progressi paragonabili a quelli di nazioni più ricche ha implicazioni politiche molto rilevanti. Questa capacità contrasta con la diffusa opinione secondo cui un paese in via di sviluppo non può affrontare spese per l'assistenza sanitaria e per l'istruzione finché non sia più solido sul piano finanziario. Si tratta di una teoria che ignora i costi relativi. L'istruzione e l'assistenza sanitaria sono attività ad alto contenuto di lavoro, come molti dei più efficienti servizi medici. In un paese con un basso costo del lavoro, questi servizi costano molto meno che in un paese più ricco. Un paese povero, quindi, pur avendo meno da spendere per questi servizi, può però spendere molto poco per realizzarli.

Non è detto che nelle nazioni ricche la speranza di vita sia superiore a quella dei paesi poveri. L'Arabia Saudita, per esempio, è ricca, ma ha una speranza di vita inferiore a quella dello Stato indiano del Kerala. Qui infatti, grazie a stanziamenti pubblici nell'istruzione, nella sanità e nell'alimentazione, nonostante il bassissimo PNL, si è potuta estendere notevolmente la speranza di vita.

Negli Stati Uniti i maschi neri di età compresa tra 35 e 54 anni hanno una probabilità di morire 1,8 volte superiore a quella dei coetanei bianchi. Le donne nere di questo gruppo di età, poi, hanno una probabilità di morte quasi tripla rispetto alle donne bianche.

ETNIA E INDICI DI MORTALITÀ NEGLI STATI UNITI

DONNE E UOMINI (ETÀ DA 35 A 54 ANNI)

BIANCHI

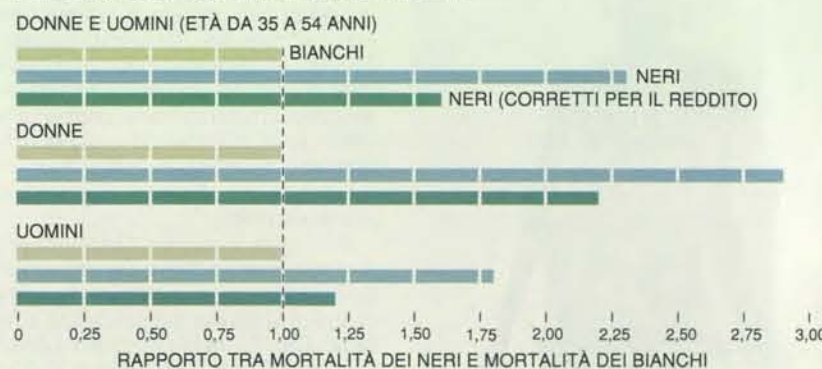
NERI

NERI (CORRETTI PER IL REDDITO)

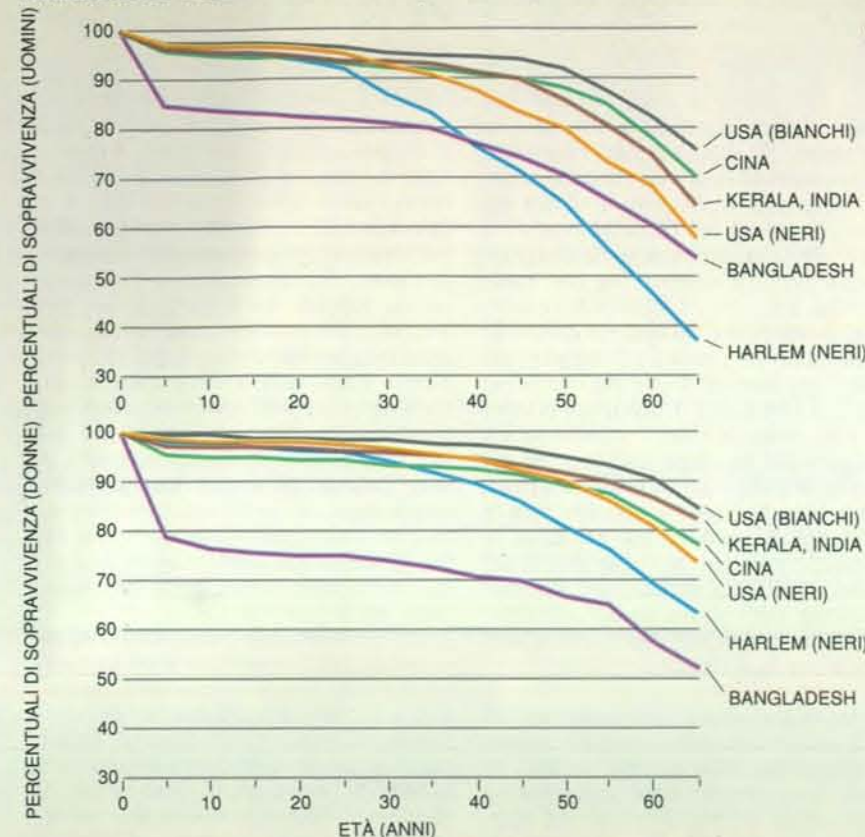
DONNE

UOMINI

0 0,25 0,50 0,75 1,00 1,25 1,50 1,75 2,00 2,25 2,50 2,75 3,00

RAPPORTO TRA MORTALITÀ DEI NERI E MORTALITÀ DEI BIANCHI

FONTE: Mac W. Otten, Jr., et al., JAMA 1990.

In Inghilterra e nel Galles, i decenni delle due guerre mondiali videro la più decisa crescita della speranza di vita, grazie alla diffusione dell'assistenza sanitaria e alla garanzia per tutti di accedere agli alimenti razionati.

FONTE: S. Preston, N. Keyfitz, R. Schoen, Causes of Death: Life Tables for National Populations, 1972.

FONTE: Per Harlem e il Bangladesh, Colin McCord e Harold Freeman, NEJM, 1990; altri dati, statistiche ufficiali sulla popolazione dagli anni ottanta. I dati sono i più recenti a disposizione.

La probabilità di sopravvivenza dell'afro-americano medio è migliore di quella di un afro-americano che viva ad Harlem, ma risulta sfavorevole se la si confronta con quella dei bianchi statunitensi e quella degli abitanti della Cina e del Kerala, paesi con redditi molto inferiori. Anche le donne nere, per quanto i dati a esse relativi siano migliori di quelli degli uomini, sono indietro rispetto alle coetanee del Kerala e della Cina.

Gli sforzi da tempo profusi nello Sri Lanka e nello Stato indiano del Kerala (che, con 29 milioni di abitanti, ha una popolazione maggiore di quella del Canada) possono servire a illustrare i meriti della spesa pubblica per l'istruzione e la sanità. Fin dagli inizi del secolo, lo Sri Lanka ha promosso programmi di alfabetizzazione e di scolarizzazione. Negli anni quaranta ha avviato una massiccia espansione dei servizi medici, e nel 1942 ha iniziato la distribuzione gratuita o sovvenzionata di riso per aumentare l'assunzione alimentare di un popolo denutrito. Nel 1940 l'indice di mortalità era del 20,6 per mille; nel 1960 era sceso all'8,6 per mille.

Cambiamenti analoghi si sono verificati nel Kerala. Benché il PNL pro capite sia notevolmente inferiore alla media dell'India, qui la speranza di vita è ora superiore a 70 anni. Questo, a fronte del bassissimo reddito, è il risultato dell'espansione dell'istruzione pubblica, dell'assistenza sociale in campo epidemiologico, dei servizi medici individuali e dei sussidi per l'alimentazione.

Questa analisi non contraddice il valido contributo che l'aumento del PNL può dare all'aumento della speranza di vita. È ovvio che la solidità economica può servire a una famiglia per ottenere una miglior nutrizione e maggiori cure mediche. Inoltre, la crescita economica può incrementare la capacità di un paese di provvedere all'istruzione, all'assistenza sanitaria e agli aiuti alimentari. Ma i risultati della crescita economica non sono sempre incanalati verso questo genere di programmi. Molte nazioni - come Arabia Saudita, Gabon, Brasile e Repubblica Sudafricana - hanno dati relativi a istruzione, sanità e previdenza decisamente peggiori di quelli di altri Stati con PNL molto inferiori, ma con una politica sociale più incisiva, come Sri Lanka, Cina, Costa Rica e Kerala. Non è pertanto necessario che i paesi poveri diventino ricchi per combattere la mortalità e aumentare la speranza di vita.

Naturalmente, il ruolo della politica sociale nel prolungamento della speranza di vita non è una peculiarità del Terzo Mondo. L'intervento pubblico nel campo della sanità, dell'istruzione e dell'alimentazione ha avuto storicamente un ruolo fondamentale nel prolungamento della vita in Occidente e in Giappone. In Inghilterra e nel Galles, i decenni antecedenti la prima e la seconda guerra mondiale furono quelli che, più di ogni altro decennio di questo secolo, videro l'aumento più significativo della speranza di vita. Gli sforzi bellici e il razionamento portarono a una più equa distribuzione del cibo, e si prestò maggiore attenzione all'assistenza medica; lo stesso National Health Service fu istituito negli anni quaranta. Eppure, in questi due decenni si ebbe la più bassa crescita del prodotto interno lordo (PIL): tra il 1911 e il 1921, essa fu addirittura negativa. L'impegno pubblico, più che il reddito individuale, fu la chiave per aumentare in quei decenni la speranza di vita.

L'analisi dei dati sulla mortalità può servire a dare una valutazione economica dei progetti sociali e della politica statale. Questa prospettiva può risultare di particolare utilità per chiarire alcuni aspetti fondamentali della disuguaglianza sociale e della povertà, e per individuare gli interventi che possono contrastarli. Uno dei problemi più impellenti che gli Stati Uniti devono affrontare è la necessità di capire più a fondo la natura della povertà economica. Il reddito è ovviamente un dato centrale per caratterizzare la povertà, ma il dibattito sulla povertà negli Stati Uniti in generale, e quella degli afro-americani in particolare, ha spesso trascurato aspetti importanti per un'eccessiva attenzione al reddito.

Come è stato più volte rilevato, due quinti dei residenti di Harlem, un quartiere del centro di New York, vivono in famiglie con livelli di reddito al di sotto della soglia nazionale di povertà. Questo fatto è sbalorditivo, ma quella soglia di povertà, per bassa che sia nel contesto statunitense, è comunque di molte vol-

te superiore al reddito medio, per esempio, di una famiglia del Bangladesh, anche tenuto conto delle differenze di prezzi e di potere d'acquisto. In un certo senso, le statistiche sulla mortalità possono risultare più efficaci per confrontare la povertà di Harlem e quella del Bangladesh. Colin McCord e Harold P. Freeman, della Columbia University e dell'Harlem Hospital, hanno già rilevato che i neri di Harlem hanno minori probabilità dei maschi del Bangladesh di raggiungere i 65 anni d'età. In realtà, dall'età di quarant'anni, gli individui di sesso maschile di Harlem si trovano in svantaggio in termini di sopravvivenza rispetto a quelli del Bangladesh.

Per rendere più significativi questi confronti, possiamo prendere in considerazione le situazioni della Cina e del Kerala, economie povere che più del Bangladesh si sono impegnate in programmi pubblici nel campo dell'assistenza sanitaria e dell'istruzione. Pur avendo una maggiore mortalità infantile, la Cina e il Kerala hanno probabilità di sopravvivenza migliori di Harlem per quanto riguarda gli individui di sesso maschile dall'adolescenza in avanti. In parte, la maggiore mortalità degli uomini di Harlem è conseguenza delle morti violente. La violenza è ad Harlem un aspetto significativo della povertà, anche se non è l'unica causa dell'alta mortalità. Rispettivamente dall'età di 35 e 30 anni la probabilità di sopravvivenza delle donne di Harlem è inferiore a quella delle donne cinesi e del Kerala.

Per di più, un problema analogo colpisce tutti gli afro-americani in generale. Anche in questo caso, i neri statunitensi hanno una mortalità infantile inferiore a quella della Cina e del Kerala. Ma se si sale la scala delle età, le donne e gli uomini neri si trovano indietro, in termini di sopravvivenza, rispetto alle donne e agli uomini della Cina e del Kerala. Non è possibile avere un'adeguata comprensione della natura e dell'ampiezza della indigenza degli afro-americani se si prende come misura il metro del reddito. Secondo questa scala, gli afro-americani sono poveri in confronto ai bianchi degli Stati Uniti, ma sono immensamente più ricchi dei cinesi e degli abitanti del Kerala. D'altra parte, in termini di vita e di morte, gli afro-americani hanno minori probabilità di arrivare a età avanzate rispetto alle popolazioni di alcuni dei paesi più poveri del Terzo Mondo.

Un'altra caratteristica della disuguaglianza razziale rivelata dai dati sulla mortalità è la povertà relativa delle donne afro-americane. In qualche modo, le donne sono in condizioni peggiori rispetto agli uomini. Nelle età comprese tra i 35 e i 54 anni, i dislivelli tra la mortalità dei bianchi e dei neri sono molto maggiori per le donne che non per gli uomini. In certa misura, le differenze tra neri e bianchi hanno a che fare con le differenze di reddito; ma anche se si operano correzioni per tener conto dei redditi, il divario rimane, almeno parzialmente. Per le donne nere, la differenza nella mortalità non può essere affatto attribuita alla diversità di reddito.

I dati sulla mortalità possono servire anche per studiare una manifestazione elementare delle discriminazioni tra sessi. Una sorprendente caratteristica demografica del mondo attuale è l'enorme variabilità geografica del rapporto tra maschi e femmine. Stando ai dati medici, le femmine, a parità di cure, tendono ad avere una mortalità inferiore a quella dei maschi. Perfino nell'utero, i feti femminili sono meno soggetti ad aborto spontaneo. Anche se i maschi sono più numerosi delle femmine alla nascita e al concepimento, in Europa e nel Nord America le donne superano gli uomini di circa il 5 per cento.

In molte zone in via di sviluppo, invece, i rapporti tra donne e uomini sono assai diversi: se nell'Africa subsahariana il rapporto è 1,02, nel Nord Africa è 0,98, in Cina, nel Bangladesh e nell'Asia occidentale è 0,94, in India è 0,93 e in Pakistan è 0,91. Per farsi un'idea delle grandezze in gioco, è utile porsi domande come: «Se i paesi come la Cina avessero un rapporto tra donne e uomini pari a quello, per esempio, dell'Africa subsahariana, quante donne in più ci sarebbero?» Prendendo a riferimento il rapporto esistente nell'Africa subsahariana, come abbiamo fatto io e Jean Drèze della Delhi School of Economics, più di 100 milioni di donne risulterebbero «mancanti» nei paesi con meno femmine che maschi: 44 milioni nella sola Cina, 37 milioni in India. Usando parametri diversi, altri sono giunti a stime comprese tra 60 e 90 milioni.

Il fenomeno delle donne mancanti testimonia che in questi paesi c'è una lunga storia di maggiore mortalità femminile e di forte discriminazione nei confronti delle donne. Io e Jocelyn Kynch dell'Università di Oxford abbiamo esaminato le registrazioni ospedaliere di Bombay e abbiamo scoperto che, per essere ricoverate in ospedale, le donne dovevano essere ammalate più gravemente degli uomini. Da un altro studio che ho effettuato insieme con Sunil Sengupta, della Visva-Bharati University, emergeva, a proposito degli interventi sanitari in campo alimentare, una sistematica discriminazione a favore dei ragazzi in due villaggi del Bengala occidentale.

Nascono più maschi che femmine, ma queste ultime hanno una mortalità inferiore: tendono, quindi, a superare di numero i maschi se entrambi i sessi ricevono la stessa assistenza sanitaria. In Europa e nel Nord America, il rapporto tra donne e uomini è di circa 1,05, anche se questa cifra risente della perdita di uomini nelle ultime guerre. In altri paesi, le donne non hanno la stessa possibilità di accesso all'assistenza sanitaria.

Pur essendoci, dietro queste discriminazioni, fattori storici e culturali, anche le istituzioni economiche sono parte in causa. I dati dimostrano che la possibilità di avere un reddito e di svolgere un lavoro fuori di casa, soprattutto se lavoro qualificato, migliora lo *status* sociale delle donne e influisce sull'attenzione che esse ricevono all'interno della famiglia. Lavorare fuori di casa, inoltre, dà alle donne l'occasione di confrontarsi con il mondo e, a volte, di interrogarsi sulla giustizia dell'ordine sociale ed economico dominante. L'alfabetizzazione, l'istruzione, la possibilità di possedere terre e di ereditare sono altre vie per migliorare la condizione delle donne.

Nel Kerala, l'economia è servita a rendere migliore la posizione delle donne. Non solo l'amministrazione statale ha una forte percentuale di donne in occupazioni di prestigio, ma, come già detto in precedenza, ha un avanzato sistema scolastico, con alti indici di alfabetizzazione per entrambi i sessi, una rete diffusa di servizi sanitari e, per un ampio e influente segmento della popolazione, una tradizione di successione matrilineare. Il rapporto femmine-maschi è attualmente pari a circa 1,04 (la cifra sarebbe un poco inferiore prendendo in considerazione gli uomini che lavorano fuori dei confini dello Stato). La speranza di vita alla nascita è nel Kerala di 73,0 anni per le donne e di 67,5 anni per gli uomini.

Gli afro-americani che vivono in ambienti urbani simili a questo hanno minori probabilità di sopravvivenza rispetto ai cittadini del Kerala. Questo divario mette in evidenza quanto siano carenti, negli Stati Uniti, le iniziative volte a dare equa soluzione ai problemi di istruzione, assistenza sanitaria, alimentazione e pace sociale.

Quella speranza media di vita è più o meno la stessa che si ha in Cina, ma la condizione delle donne è relativamente migliore nel Kerala. Il Governo cinese ha lottato per sradicare la disuguaglianza tra i sessi, e la Cina ha un elevato livello di occupazione femminile. L'analfabetismo tra le donne, però, è molto maggiore che nel Kerala. In Cina l'elevata mortalità infantile delle femmine può essere in parte collegata all'effetto delle misure obbligatorie di controllo delle nascite - la parziale imposizione della cosiddetta politica del figlio unico - in una società in cui domina il privilegio maschile.

Questo articolo non si occupa direttamente della fertilità e della pianificazione familiare, ma vorrei rilevare come il controllo coatto delle nascite sia pericoloso per quanto riguarda la discriminazione sessuale. Ottimi argomenti contro questo obbligo si fondano innanzitutto su considerazioni di libertà e indipendenza. Ma il possibile effetto di questa misura sulla mortalità femminile aggiunge un'altra dimensione al dibattito. Spesso, nelle discussioni sulla necessità di una forzosa pianificazione familiare nel Terzo Mondo, si cita il successo ottenuto in Cina nel rallentamento della natalità. È vero che l'indice di natalità in Cina, pari al 21 per mille, appare molto favorevole se lo si confronta con il 30 per mille dell'India (e la media del 38 per mille di paesi a basso reddito diversi dalla Cina e dall'India). Ma l'indice di natalità nel Kerala, 20 per mille, è comparabile al 21 per mille della Cina, senza che sia stata attuata alcuna politica di controllo demografico e senza il problema della mortalità infantile femminile.

Una notevole mole di dati demografici dimostra che alla diminuzione degli indici di mortalità segue molto spesso una riduzione degli indici di natalità. Questo fenomeno dipende dalla minore necessità di procreare numerosi figli per assicurare la sopravvivenza di alcuni, e testimonia l'interdipendenza tra controllo delle nascite e controllo delle morti: dare accesso alla contraccezione può associarsi efficacemente all'erogazione di assistenza sanitaria. In parallelo alla diminuzione dell'indice di mortalità, nel Kerala si è verificata anche una riduzione dell'indice di natalità: dal 44 per mille nel periodo tra il 1951 e il 1961, al 20 per mille tra il 1988 e il 1990.

I dati sulla mortalità costituiscono un indicatore della povertà economica che va molto al di là del reddito e delle disponibilità finanziarie. Valutare i risultati economici in termini di vita e di morte può portare l'attenzione su problemi pressanti di politica economica. È un punto di vista che può servire a migliorare la comprensione della denutrizione, dell'assistenza sanitaria e della disuguaglianza tra i sessi, così come della povertà e della disuguaglianza razziale, anche nelle nazioni ricche del mondo occidentale. La necessità di ampliare gli orizzonti dell'economia tradizionale, includendo l'economia della vita e della morte, è altrettanto acuta in Occidente che nell'Africa subsahariana colpita da carestie e fame.

BIBLIOGRAFIA

SEN AMARTYA, *Poverty and Famines: An Essay on Entitlement and Deprivation*, Oxford University Press, 1981.

CALDWELL JOHN C., *Routes to Low Mortality in Poor Countries* in «Population and Development Review», 12, n. 2, giugno 1986.

DRÈZE JEAN e SEN AMARTYA, *Hunger and Public Action*, Oxford University Press, 1989.

OTTEN MAC W., Jr., TEUTSCH STEVEN M., WILLIAMSON DAVID F. e MARKS JAMES S., *The Effect of Known Risk Factors on the Excess Mortality of Black Adults in the United States* in «Journal of the American Medical Association», 263, n. 6, 9 febbraio 1990.

SEN AMARTYA, *Inequality Reexamined*, Harvard University Press, 1992.

ANAND SUDHIR e RAVALLION MARTIN, *Human Development in Poor Countries: On the Role of Private Incomes and Public Services* in «Journal of Economic Perspectives», 7, n. 1, inverno 1993.

Il confine tra nucleo e mantello

Questa remota zona di interazione, finalmente «accessibile» grazie alla alla sismologia e a sofisticate tecniche di simulazione in laboratorio, va rivelandosi come la regione più dinamica in assoluto del nostro pianeta

di Raymond Jeanloz e Thorne Lay

Circa 2900 chilometri ci separano dalla più straordinaria struttura interna del pianeta Terra: se ci si potesse andare in automobile, occorrerebbero poco meno di tre giorni. Per lo più ignorata in passato dalla ricerca, la remota regione tra il mantello inferiore e il nucleo esterno si sta dimostrando essenziale per la comprensione dell'evoluzione chimica e termica del nostro pianeta. Non più considerata semplicemente come una superficie che delinea il confine tra il mantello, roccioso, e il nucleo esterno, composto di ferro liquido, la regione tra nucleo e mantello potrebbe rivelarsi come la parte del pianeta più attiva dal punto di vista geologico. Sembra che le sue caratteristiche siano cambiate completamente nel corso della storia terrestre, e che le sue proprietà fisiche ora manifeste varino da punto a punto in prossimità del limite inferiore del mantello. Le variazioni delle proprietà fisiche sui due lati dell'interfaccia tra nucleo e mantello sono infatti addirittura superiori a quelle che si hanno in corrispondenza della superficie terrestre, alla separazione fra atmosfera e superficie rocciosa.

Si ritiene che la notevole eterogeneità della regione di confine tra nucleo e mantello influenzi molti processi geologici a scala globale (si veda l'articolo *Il mantello della Terra* di D. P. McKenzie in «Le Scienze» n. 183, novembre 1983). La dinamica di quella zona provoca una piccola alterazione nell'inclinazione dell'asse di rotazione terrestre e nelle caratteristiche del campo geomagnetico. Le variazioni che avvengono nella regione nucleo-mantello influenzano anche i fenomeni di convezione nel mantello terrestre, che sono responsabili dei movimenti dei continenti e delle zolle litosferiche.

Il primo indizio del fatto che qualcosa di inusuale stava accadendo alle profondità alle quali nucleo e mantello si incontrano, si manifestò a metà degli anni trenta. A fornire tale indizio furono le onde generate dai terremoti: all'interno di gran parte del mantello, la velocità delle onde sismiche aumenta generalmente con la profondità e presenta trascurabili variazioni laterali. Questo può significare che all'aumentare della profondità la Terra diventa sempre più «semplice»: manifesta cioè una sempre maggiore uniformità nella sua composizione e struttura. La grande varietà di strutture geologiche e di rocce osservate appena al di sotto della superficie mostra come quest'ultima sia invece la regione più complessa.

Peraltro questo comportamento della velocità delle onde sismiche si mantiene solo fino a un certo punto. Nelle ultime centinaia di chilometri prima di giungere al centro della Terra, nella zona più profonda del mantello, appena prima che appaia il nucleo, la velocità media delle onde sismiche non aumenta sensibilmente e si manifestano cambiamenti più significativi di velocità tra regione e regione (*si veda l'illustrazione alle pagine 26 e 27*). Questo effetto è quasi impercettibile: corrisponde a una differenza di pochi punti percentuali. D'altra parte, secondo gli standard geologici, questa piccola differenza corrisponde a enormi variazioni nella struttura e/o nella temperatura. I primi studiosi riconobbero l'importanza di queste variazioni semplicemente esaminandole nel sovrastante mantello inferiore, e conseguentemente chiamarono questa regione, che ha uno spessore compreso circa tra 200 e 400 chilometri, strato D''.

L'origine di questo nome è essenzialmente storica. I primi geologi avevano contrassegnato i vari strati del pianeta con lettere dell'alfabeto, e non con i nomi di «crosta», «mantello» e «nucleo». Ogni suddivisione secondaria identificata in seguito doveva quindi essere contrassegnata con apici associati alla lettera. Anche se successivamente i nomi dei vari strati sono cambiati, lo strato D'' ha conservato la sua denominazione originaria.

Numerose interpretazioni sono state proposte per spiegare le proprietà sismiche dello strato D''. Sfortunatamente, troppe erano le possibili spiegazioni e troppo scarsa l'informazione che potesse consentire di definire in modo univoco le caratteristiche di quello strato. Per disporre di migliori descrizioni dello strato D'' si dovette attendere il progresso tecnologico degli anni ottanta quando, utilizzando schiere di strumenti installati su tutto il pianeta, divenne finalmente possibile raccogliere e analizzare una quantità di dati sufficiente per ricavare immagini tridimensionali dell'interno della Terra (si veda l'articolo *La tomografia sismica* di Don L. Anderson e Adam M. Dziewonski in «Le Scienze» n. 196, dicembre 1984). I sismometri usati operano fondamentalmente nell'intervallo tra 1 e 0,0003 hertz (queste frequenze acustiche sono ben al di sotto dell'intervallo di udibilità per l'orecchio umano, che si estende circa da 20 a 20 000 hertz). La tomografia sismica viene sovente paragonata con la tomografia assiale computerizzata, o TAC, che viene utilizzata in diagnostica medica: in realtà, dato che essa utilizza le onde acustiche, è più correttamente paragonabile alle ecografie che si effettuano, per esempio, per seguire il decorso di una gravidanza. Il principale limite della tomografia sismica risiede nella sua scarsa risoluzione: immagini di dettagli di dimensioni inferiori ai 2000 chilometri non vengono generalmente rilevate.

Malgrado ciò, la tomografia sismica ha contribuito a quantificare le proprietà dello strato D'', mostrando come quella regione si differenzi assai nettamente dal sovrastante mantello. Il fatto che le variazioni nella velocità delle onde sismiche si manifestino su aree di dimensione continentale dimostra che lo strato D'' è in prevalenza caratterizzato da strutture di grande scala. Peraltro, la tomografia sismica non è valsa a spiegare le cause di questa variabilità nelle proprietà fisiche. Potevano esistere alla base del mantello grandi strutture chimicamente differenziate, o forse le eterogeneità indicavano semplicemente differenze di temperatura?

Per rispondere a queste domande, uno di noi (Lay) ha iniziato nei primi anni ottanta a sviluppare un nuovo metodo per esplorare il confine tra nucleo e mantello. L'idea era quella di utilizzare il calcolatore per analizzare tutte le caratteristiche del fronte d'onda sismico osservato e non solo la velocità delle onde, come avveniva nella tomografia sismica. Questo tipo di analisi della forma d'onda è un metodo molto efficace: è infatti in grado di distinguere strutture con una dimensione di poche decine di chilometri. Si ottiene in tal modo una risoluzione di gran lunga superiore rispetto ai 2000 chilometri raggiunti dalla tomografia sismica. Lo svantaggio è costituito dal fatto di potere osservare solamente parti limitate del confine tra nucleo e mantello: non ci sono infatti sufficienti terremoti o altre sorgenti di energia sismica per poter ottenere una immagine

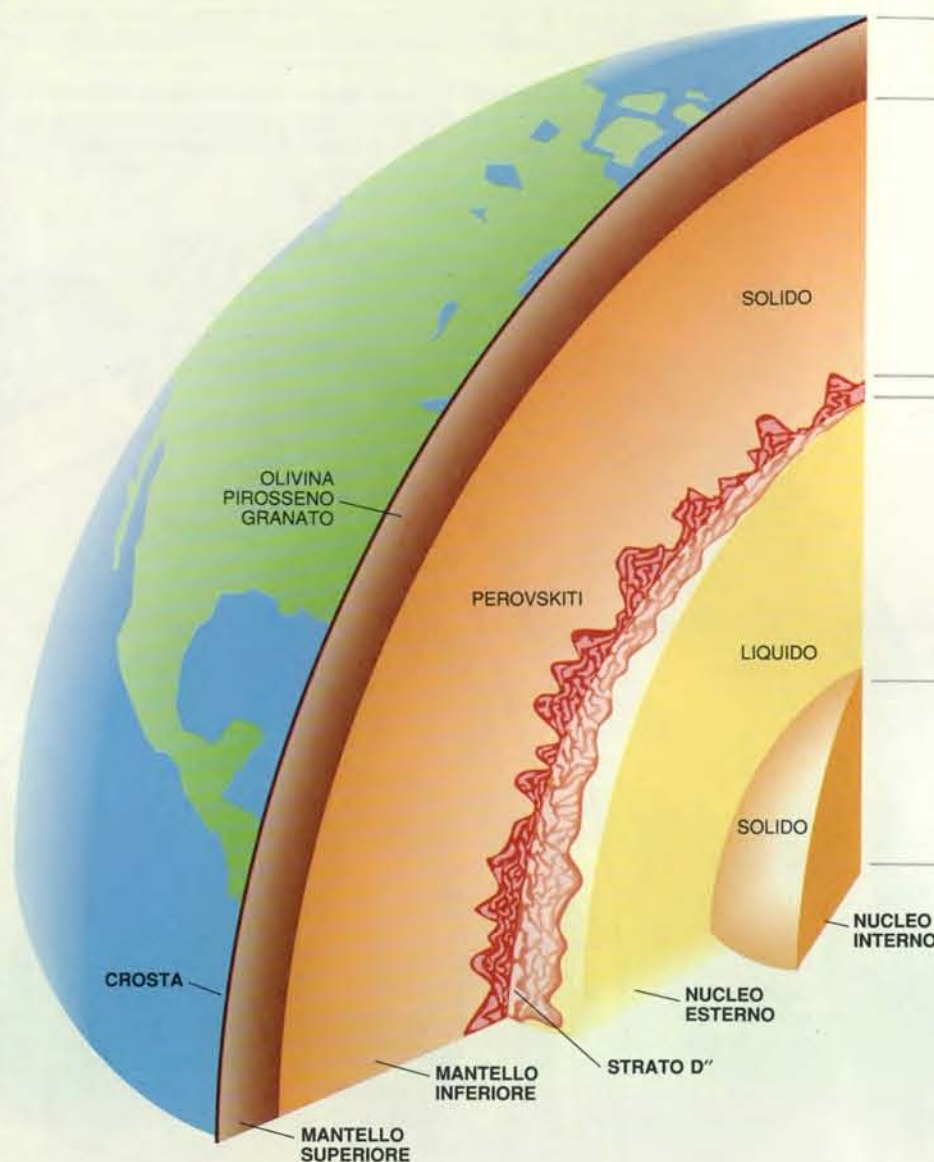
La velocità delle onde sismiche varia da una regione all'altra dell'interno della Terra, come illustra questa immagine ottenuta con la tomografia sismica. In alcune regioni (*in blu*) le onde si muovono più velocemente di quanto non accada in media a quella profondità; in altre (*in giallo*) le onde sono più lente della media. Tali variazioni possono far pensare a differenze nella composizione del mezzo. Nell'immagine non è evidenziata la complessità del confine tra nucleo e mantello (la superficie esposta del nucleo esterno), in quanto le eterogeneità sono troppo esigue per il potere di risoluzione del metodo.

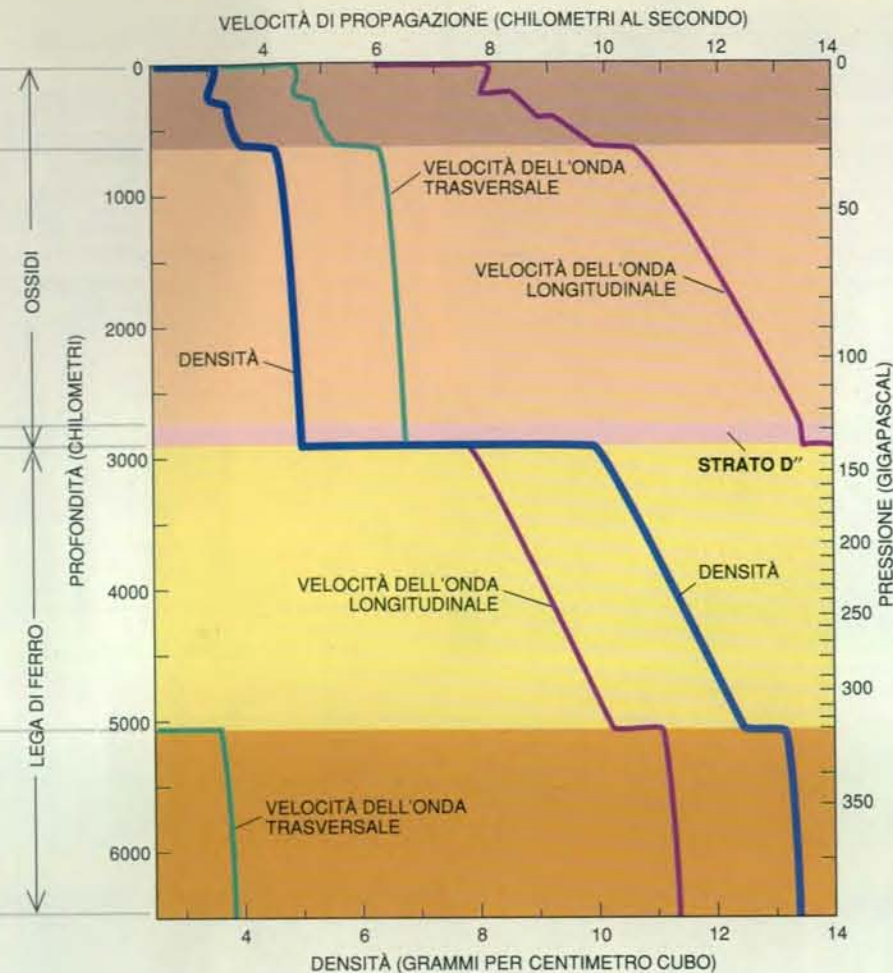
globale a un tale livello di dettaglio.

Gli studi della forma d'onda suggeriscono che regioni situate all'interno dello strato D″ siano più differenziate di quanto si ritenesse in passato. Numerosi gruppi di ricerca che studiano il confine tra nucleo e mantello al di sotto della Siberia settentrionale hanno per esempio accertato che la velocità delle onde acustiche varia notevolmente su piccole distanze, tanto da far rilevare forme d'onda differenti a sismometri situati a poca distanza l'uno dall'altro. La miglior spiegazione che si può fornire per questa scoperta scaturisce dall'ipotesi che la eterogeneità nelle velocità delle onde sismiche sia quantitativamente elevata e che si manifesti su distanze inferiori alla risoluzione disponibile, pari ad alcune decine di chilometri. Studiando le forme d'onda è anche possibile elaborare una mappa delle differenze di spessore nello strato D″. In molti luoghi la sommità dello strato D″ causa un repentino aumento nella velocità delle onde, un processo che porta alla riflessione dell'energia sismica. Ciò ha rivelato che lo spessore dello strato D″ varia notevolmente: lo strato può essere così sottile da non essere rilevabile, oppure può avere addirittura uno spessore di 300 chilometri.

Il gruppo di Stanley M. Flatté dell'Università della California a Santa Cruz ha contribuito a confermare la grande variabilità dello strato D″. Durante la seconda metà degli anni ottanta, assieme ai suoi colleghi ha iniziato ad applicare i nuovi metodi di analisi delle onde ai segnali ottenuti dalle onde sismiche diffuse nel mantello profondo. Il loro metodo si basa su una descrizione statistica di come le onde si propagano attraverso una sostanza che tende a diffonderle, proprio come la nebbia e le nubi fanno con le onde luminose. Il metodo di Flatté consiste nell'osservare come il fronte d'onda di un terremoto cambi forma dopo aver viaggiato attraverso la regione D″. Un terremoto emette inizialmente un'onda sferica regolare, ma non appena quell'onda viene riflattata e quindi diffusa da variazioni di proprietà sismiche, come le forti eterogeneità al confine tra nucleo e mantello, il fronte d'onda risulta deformato (*si veda l'illustrazione a pagina 29*).

Per riuscire a misurare il grado di «corrugamento» del fronte d'onda, occorre una fitta schiera di sismometri. Grazie a una simile schiera collocata in Norvegia, Flatté ha potuto mostrare come la regione D″ appaia alle onde sismiche priva di un contorno nitido: è come se la sua immagine apparisse sfocata; le dimensioni delle eterogeneità che la caratterizzano sono dell'ordine di appena 10 chilometri. Le osservazioni sismologiche indicano così che la regione D″ è uno strato eterogeneo il cui spessore presenta variazioni laterali.

Al contrario dello strato D″, il confine tra nucleo e mantello (sul quale lo strato D″ appoggia) appare regolare e netto. Nel 1992 John E. Vidale e Harley Benz dello US Geological Survey hanno dimostrato brillantemente la presenza di questa interfaccia, utilizzando un gran numero di stazioni di rilevamento sismico distribuite negli Stati Uniti occidentali. La schiera di sismometri in genere rileva i terremoti a livello regionale, ma Vidale e Benz l'hanno utilizzata per trovare onde sismiche riflesse dal confine tra nucleo e mantello. È significativo che le onde sismiche arrivassero in modo coerente in oltre 900 stazioni della schiera: ciò implica che il confine tra nucleo e mantello rappresenta, per lo meno nell'area misurata, una brusca transizione da mantello a nucleo, che riflette ben il 50 per cento delle onde sismiche e trasmette il rimanente. Le analisi delle onde riflesse e trasmesse mostrano che la profondità di questo confine varia al più di qualche chilometro.

Gli studi sulle onde sismiche hanno fatto molto per chiarire la natura dello strato D″ e del confine tra nucleo e mantello; però la inaccessibilità di tali regioni ha impedito ai geofisici di giungere a una piena comprensione sul come sia avvenuta la formazione di strutture tanto complicate.

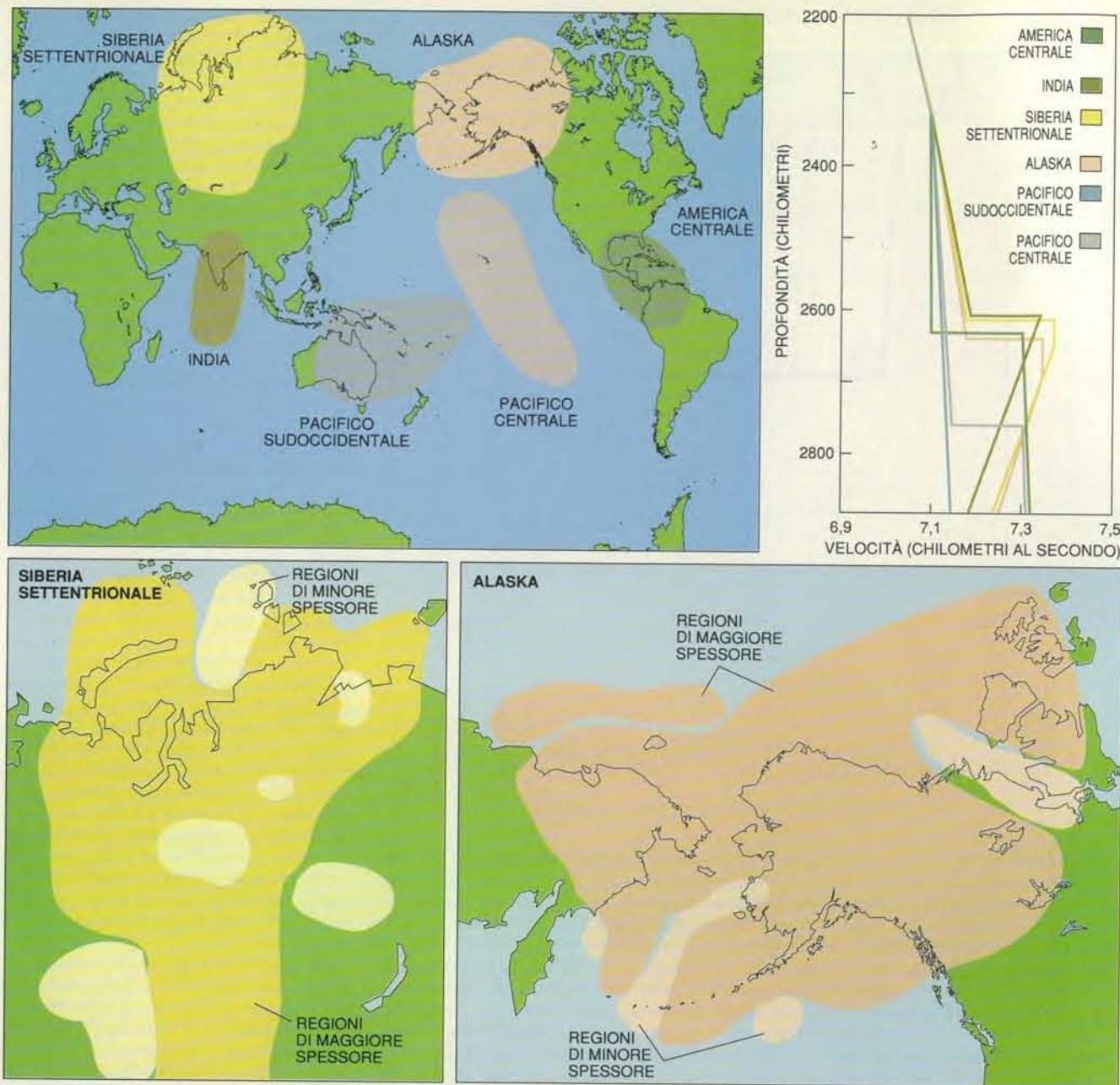
Se gli studi sismici non sono in grado di penetrare sino alle remote profondità della Terra, perché non portarsi il nucleo e il mantello in superficie? Questo è precisamente il metodo seguito da molti ricercatori, tra i quali uno di noi (Jeanloz). In particolare, noi abbiamo cercato di riprodurre le elevate pressioni e temperature che esistono nel mantello profondo e nel nucleo, grazie alla disponibilità di un'avanzata tecnica di laboratorio con la quale si riesce a comprimere minuscoli campioni tra due punte di diamante e a scaldarli per mezzo di un raggio laser di alta potenza (si veda l'articolo *Incudini di diamante per alte pressioni* di A. Jayaraman in «Le Scienze» n. 190, giugno 1984). Verso il 1986 le celle di diamante erano ormai in grado di generare pressioni maggiori di quelle esistenti al centro della Terra.

Lo spaccato della Terra nella pagina a fronte mostra le principali regioni del pianeta. La crosta e il mantello superiore consistono di ossidi come olivina, pirosseno e granato e il mantello inferiore di silicati con struttura di tipo della perovskite. Il nucleo è costituito da una lega di ferro, liquida nella parte esterna e solida nel centro. Gli strati corrispondono alle variazioni osservate nella densità e nella velocità delle onde sismiche, mentre si propagano attraverso la Terra. Sia la densità sia la velocità delle onde aumentano in funzione della profondità eccetto che nello strato D″. L'energia sismica può propagarsi sotto forma di onde trasversali (onde che oscillano perpendicolarmente alla direzione di propagazione) e sotto forma di onde di compressione, o longitudinali (che oscillano lungo la direzione di propagazione). Le onde trasversali non possono propagarsi nel nucleo esterno, essendo questo allo stato liquido. Tali onde ricompaiono nel nucleo interno in quanto una frazione delle onde di compressione si trasforma in onde trasversali alla interfaccia liquido-solido.

La durezza del diamante non è la sola ragione per usarlo come materiale per l'incudine: la sua utilità risiede anche nella trasparenza. Il raggio laser può essere focalizzato direttamente attraverso il diamante per riscaldare il campione sino a migliaia di gradi Celsius ed è inoltre possibile osservare il campione mentre viene sottoposto a pressioni e temperature estremamente elevate. La temperatura del campione viene determinata misurando la radiazione termica che esso emette attraverso il diamante. In questo modo, si può valutare il punto in cui il materiale raggiunge il «calor rosso» oppure il «calor bianco», allo stesso modo in cui gli astronomi stimano la temperatura superficiale delle stelle. Utilizzando la cella di diamante riscaldata con il laser, noi possiamo simulare le stesse temperature e pressioni esistenti in prossimità del confine tra nucleo e mantello. Volevamo osservare che cosa sarebbe accaduto qualora avessimo posto il materiale che costituisce il nucleo esterno in contatto con minerali del mantello inferiore.

Evidentemente dovevamo essere a conoscenza della reale natura di questi materiali costitutivi. Per identificare il materiale del mantello, Elise Knittle, insieme con Jeanloz, ha proseguito una ricerca condotta precedentemente da gruppi della Australian National University, della Carnegie Institution di Washington e di altri centri. Ci siamo basati su precedenti esperimenti, su modelli teorici e sul fatto che la pressione nel mantello inferiore supera i 20 gigapascal (200 000 volte la pressione atmosferica).

Da quelle informazioni abbiamo dedotto che una sola fase mineralogica di alta pressione deve dominare il mantello inferiore. Questo minerale è una forma densa di silicato di ferro e magnesio, ovvero $(Mg, Fe)SiO_3$, un composto compatto e chimicamente semplice che può formarsi solo a pressioni superiori a 20 gigapascal. Poiché esso ha la stessa struttura cristallina della perovskite ($CaTiO_3$), esso viene per questa ragione chiamato perovskite a silicato di magnesio. La roccia del mantello inferiore probabilmente contiene anche minori quantità di magnesiowüstite, una combinazione di ossido di magnesio (MgO) e wüstite (FeO). Questa composizione è completamente diversa dalla natura delle rocce che si trovano in prossimità della superficie della Terra o su di essa: questi tipi di roccia sono composti di molti minerali differenti e complessi, che reagiscono chimicamente e si trasformano in nuovi minerali quando sono soggetti a modeste variazioni di pressione o di temperatura. La semplicità della composizione chimica del mantello inferiore, così dedotta, si accorda bene con i dati ricavati dalle onde sismiche, i quali mostrano come quest'ultimo sia relativamente privo di strutture (eccezion fatta per lo strato D″). Questa corrispondenza ci rassicura sulla correttezza delle nostre deduzioni in merito ai materiali da scegliere per le nostre simulazioni di laboratorio.

La determinazione della composizione del nucleo è stata più semplice. Gli studi sismologici condotti oltre 50 anni fa hanno consentito ai geofisici di dedurne la struttura: il nucleo consiste di un centro solido circondato da materiale fuso di natura metallica, più precisamente una lega di ferro. È la rotazione interna di questo strato di ferro fuso a generare il campo magnetico terrestre.

Dopo aver identificato i composti coinvolti nei fenomeni studiati, Elise Knittle ha condotto una serie di esperimenti nei quali il ferro liquido veniva messo in contatto ad alte pressioni con perovskite, e ha potuto osservare che la perovskite reagisce energicamente con il ferro liquido anche se queste sostanze restano in contatto solo per pochi secondi. La natura di questa reazione è piuttosto interessante ed inaspettata. I suoi prodotti sono un miscuglio di ossidi elettricamente isolanti (perovskite a silicato di magnesio e stishovite), siliciuro di ferro (FeSi) e wüstite. Si riteneva che la wüstite non fosse in grado di formare leghe metalliche ad alcuna temperatura o pressione. Da un punto di vista qualita-

La velocità delle onde trasversali nello strato D″ è diversa a seconda della zona, come si vede dalle sei regioni (*aree in colore, in alto a sinistra*) studiate più a fondo. La corrispondente distribuzione della velocità in funzione della profondità (*in alto a destra*) mostra come tutte le regioni presentino una discontinuità in corrispondenza dello strato D″. Dal grafico si può notare inoltre come la profondità dello strato D″ vari da una regione all'altra. Le mappe ingrandite (*in basso*) per aree al di sotto della Siberia settentrionale e dell'Alaska riassumono l'eterogeneità dello strato D″ mostrando come vi sia un'alternanza di regioni di grande spessore (*aree scure*) e di parti così sottili da risultare sismicamente invisibili (*aree chiare*).

tivo, la wüstite può reagire a questo modo perché il suo atomo di ossigeno assume ad alte pressioni le caratteristiche chimiche normalmente attribuite al suo vicino nel sistema periodico degli elementi, lo zolfo. Solfuri metallici come il bisolfuro di ferro (pirite) sono naturalmente ben conosciuti.

Gli esperimenti hanno inoltre mostrato che il ferro liquido inizia a reagire con le sostanze del mantello a pressioni comprese tra 20 e 30 gigapascal, molto inferiori a quelle che si verificano al confine tra nucleo e mantello (136 gigapascal). Per questa ragione, le reazioni hanno probabilmente continuato a svolgersi sin dalle prime fasi della storia del pianeta, quando cioè la Terra stava ancora sviluppandosi, e il nucleo in formazione avrebbe potuto trovarsi a pressioni inferiori a 136 gigapascal. È probabile che tali reazioni chimiche abbiano significativamente alterato il sistema nucleo-mantello. Probabilmente un quantitativo considerevole di ossigeno è stato trascinato durante la storia geologica nel metallo del nucleo, o si è legato con esso; ciò ha provocato il lento dissolvimento della roccia del mantello inferiore nel metallo liquido del nucleo esterno, un processo tuttora in corso. Berni J. Alder del Lawrence Livermore National Laboratory lo aveva suggerito più di 25 anni fa e i nostri esperimenti convalidano la sua congettura.

Una delle importanti conseguenze di questa ipotesi è che essa offre una semplice spiegazione della ragione per la quale le proprietà del nucleo sono simili

ma non esattamente uguali a quelle del ferro a pressione e temperatura corrispondenti. Ancor più degno di nota è il fatto che la densità del nucleo esterno è circa il 10 per cento più bassa di quella del ferro puro (si veda l'articolo *Il nucleo della Terra* di Raymond Jeanloz in «Le Scienze» n. 183, novembre 1983). Ma, come indicato dalla ipotesi di Alder e dai nostri esperimenti con la cella di diamante, il nucleo non può essere costituito interamente di ferro, in quanto un nucleo di solo ferro sarebbe stato contaminato da reazioni con la roccia sovrastante nel corso dei tempi geologici. È abbastanza plausibile che il nucleo non sia mai stato composto di ferro puro; al contrario esso probabilmente conteneva anche un po' di nichel, zolfo e altri costituenti secondari: tale ipotesi si fonda sul ritrovamento di meteoriti ricche di ferro, considerate resti parziali dei materiali a partire dai quali la Terra prese forma, e che hanno una composizione simile. Come il ferro puro, queste leghe ricche di ferro possono reagire con composti rocciosi in condizioni di alte pressioni e temperature, formando una lega con l'ossigeno.

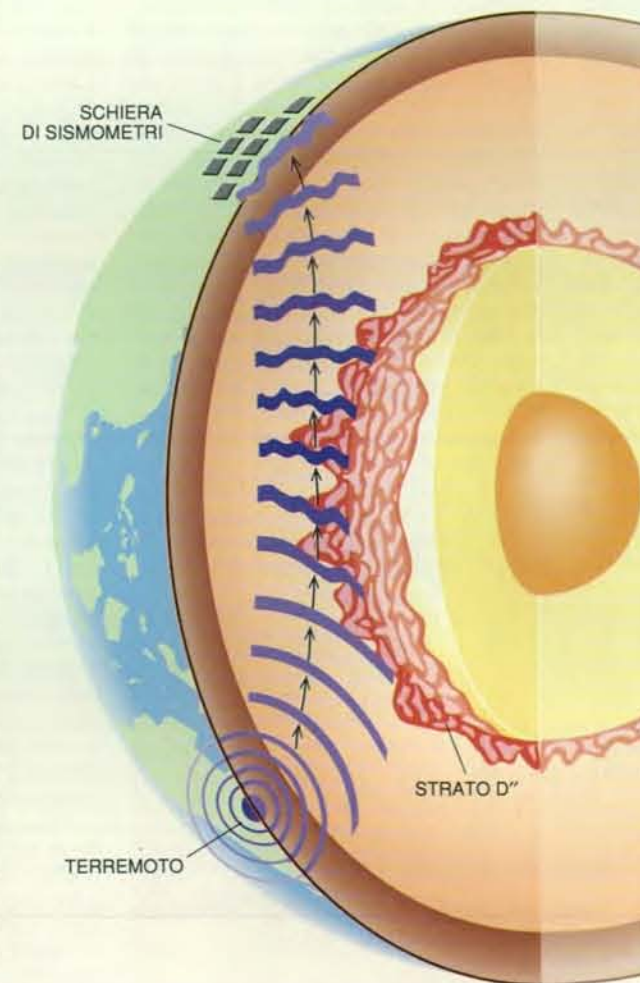
Sulla base dei nostri esperimenti, il liquido denso del nucleo esterno si infiltra nella roccia, probabilmente per capillarità: il metallo fuso penetrerebbe attraverso i minuscoli interstizi tra i grani di minerale che sono alla base del mantello. Si valuta che per effetto della capillarità il liquido del nucleo possa spingersi verso l'alto per alcune decine o centinaia di metri al di sopra del confine tra nucleo e mantello. La reazione tra il liquido del nucleo e la roccia del mantello probabilmente si completa in meno di un milione di anni, che alla scala dei tempi geologici è come dire istantaneamente.

Il liquido, tuttavia, non deve necessariamente muoversi sempre verso l'alto e vincere la gravità: essendo probabile che l'interfaccia tra il mantello e il nucleo non sia perfettamente piana, il fluido metallico potrebbe infiltrarsi lateralmente e verso il basso all'interno della roccia del mantello, da regioni nelle quali il confine tra nucleo e mantello dista maggiormente dal centro della Terra.

I dati degli studi geodetici e sismologici indicano che la topografia del confine tra nucleo e mantello presenta deviazioni dalla perfetta planarità; la misura di queste irregolarità varia tra qualche centinaio di metri e alcuni chilometri. Per questo motivo, la zona di infiltrazione e di reazione chimica diretta tra il liquido del nucleo e la roccia del mantello ha uno spessore non superiore a poche centinaia o al massimo migliaia di metri. Questa stima spiega perché gli studi sulle onde sismiche di solito non rilevano segnali che denuncino il verificarsi di reazioni al confine tra nucleo e mantello: lo spessore della zona di reazione è infatti inferiore alle tipiche lunghezze d'onda sismiche. Inoltre, in un qualsiasi momento, solo una modesta frazione della zona di reazione consiste di liquido. Pertanto, la presenza di una piccola quantità di materiale allo stato fluido non altererebbe in modo percettibile la velocità delle onde sismiche nel mantello inferiore.

In che modo queste reazioni chimiche al confine tra nucleo e mantello valgono a spiegare le caratteristiche dello strato D″ osservate? La risposta va trovata in un complesso processo che dipende indirettamente dalle forze agenti sull'interfaccia nucleo-mantello. Queste ultime derivano dalla energia termica del nucleo sottostante, che riscalda la roccia alla base del mantello. Conseguentemente, la parte riscaldata del mantello si muove verso l'alto in un periodo di decine o centinaia di milioni di anni, tempo assai più lungo di quello necessario per la reazione tra nucleo e mantello, che ha luogo in meno di un milione di anni. La convezione scompiglia la zona di reazione al confine tra nucleo e mantello, trascinandola verso l'alto ed esponendo roccia intatta del mantello al fluido corrosivo del nucleo: si tratta della stessa forza che, in superficie, spinge le zolle tettoniche a muoversi.

La convezione del mantello non trascina i fluidi molto lontano; il metallo liquido, all'inizio del suo movimento verso l'alto, con tutta probabilità defluisce dalla roccia porosa come da una spugna. I prodotti cristallini arricchiti in ferro che provengono dalla zona di reazione, come la wüstite, sono con facilità incorporati nel flusso del mantello. Il lento movimento convettivo del mantello solleva le leghe cristallizzate per una piccola distanza prima che la loro densità di solidi metallici non le costringa a sprofondare nuovamente: questi solidi essenzialmente assomigliano ai sedimenti che restano sul fondo di una tazza di caffè alla turca.

Le sostanze ricche di leghe tenderebbero così a raccogliersi al fondo del mantello, specialmente in prossimità di zone di risalita, proprio come avviene per i cumuli di neve che si formano durante una tempesta. La dispersione verso l'alto induce una infiltrazione di materiale proveniente dal nucleo e provoca la formazione di una zona di rimescolamento di maggior spessore; il mescolamento dei prodotti di reazione e del mantello non interessato dalla reazione causa la eterogeneità di comportamento sismico. Viceversa, regioni di sprofondamento provocherebbero una dispersione dei sedimenti e perciò tenderebbero a rendere più sottile lo strato D″ e a spostare più in basso il confine tra nucleo e mantello. Un modello messo a punto da Louise Kellogg dell'Università della California a Davis in collaborazione con Norman H. Sleep della Stanford University suggerisce che le le-

La distorsione di un fronte d'onda sismico consente di analizzare le eterogeneità esistenti nelle caratteristiche dello strato D″. Inizialmente le onde generate da un terremoto sono regolari. Quando esse si propagano attraverso la regione D″, i loro fronti d'onda si deformano. Questo fenomeno viene misurato da una fitta schiera di sismometri situati a grande distanza dal centro focale del terremoto. In Norvegia una schiera di questo tipo fu originariamente installata per rilevare le onde sismiche generate da esperimenti nucleari sotterranei.

ghe metalliche in aree limitate della zona di reazione possano venire trascinate verso l'alto per parecchie centinaia di chilometri all'interno del mantello. Un processo di questo tipo richiederebbe decine di migliaia di anni.

I flussi di materiali ricchi di leghe alla base del mantello risolvono un importante mistero: essi spiegherebbero la variazione nello spessore dello strato D″ osservata dai sismologi. Per di più, i calcoli effettuati indicano che la dimensione verticale dei flussi nel mantello è confrontabile con quella delle parti dello strato D″ di maggior spessore. Dati i miliardi di anni necessari per il progressivo accumulo dei sedimenti metallici, è plausibile che gran parte della complessità e delle variazioni nello spessore dello strato D″ risultino dal modo in cui il flusso del mantello modula lo strato reattivo ricco di leghe. Il flusso può anche aver trascinato nella sua scia altro materiale denso del mantello o prodotti del nucleo. Noi sospettiamo che i sedimenti di reazione si possano raccogliere, sebbene in minor misura, sul lato interno del confine tra nucleo e mantello, dove è probabile che si trovi una più sottile versione dello strato D″, appena all'interno del nucleo esterno liquido.

In virtù della intensa dinamica che ha luogo 2900 chilometri al di sotto della superficie della Terra, non ci si dovrebbe sorprendere del fatto che le forze nel sistema nucleo-mantello possano far sentire la loro presenza da un capo all'altro del pianeta. Di fatto alcuni studiosi hanno trovato elementi che suggeriscono come la zona nucleo-mantello influenzi due fenomeni osservabili in superficie: l'alterazione nell'inclinazione dell'asse di rotazione della Terra (la cosiddetta nutazione) e il campo geomagnetico.

Bruce A. Buffett, lavorando assieme a Irwin I. Shapiro alla Harvard University, ha concluso che il confine tra nucleo e mantello influenza le nutazioni della Terra. Egli è giunto ad affermarlo dopo aver svolto calcoli accurati di tale fenomeno, misurandolo per mezzo dell'interferometria a lunghissima base (VLBI), una tecnica alla quale i radioastronomi spesso si affidano per effettuare misurazioni con alta precisione di oggetti stellari. Diverse forze di tipo mareale erano state considerate come sole responsabili delle nutazioni della Terra. Tali meccanismi comprendono l'attrito dell'atmosfera e degli oceani con la superficie solida della Terra così come le interazioni gravitazionali della Terra con il Sole e la Luna. Buffett ha scoperto, tuttavia, una componente delle nutazioni che non poteva venire spiegata con le forze di marea: incoraggiato dai risultati degli esperimenti compiuti con le incudini di diamante, egli ha considerato la possibilità che una sottile zona di reazione al confine tra nucleo e mantello potesse offrire una spiegazione per la componente anomala della nutazione.

Egli ha dimostrato che una zona di reazione di questo tipo può facilmente giustificare la nutazione se il materiale contenuto nello strato è elettricamente conduttore, come gli esperimenti hanno permesso di appurare. Le linee del campo magnetico che escono dal nucleo indurrebbero il flusso di deboli correnti elettriche all'interno del materiale conduttore; queste deboli correnti a loro volta producono i loro propri campi magnetici. Questi ultimi interagiscono con le linee del campo geomagnetico principale, allo stesso modo in cui i poli di un magnete possono attrarre o respingere. In sostanza, il nucleo e il mantello si comportano come due magneti che si respingono, con un accoppiamento che influisce sulle nutazioni. I dati rilevati con la VLBI possono venire spiegati in modo convincente se si assume una zona di reazione eterogenea che contenga metallo e abbia uno spessore di alcune centinaia di metri.

Viene così confermato quanto prospettato dai nostri esperimenti, che avevano previsto per la zona di reazione una configurazione proprio come questa. Ci si aspetta che i prodotti di reazione alla base del mantello consistano in parte di leghe conduttrici di elettricità, come siliciuro di ferro e wüstite: una zona nella quale sia presente un 15-20 per cento di lega sarebbe sufficiente per spiegare le nutazioni. La nostra conclusione secondo la quale la zona di reazione avrebbe uno spessore di centinaia di metri e fluttuerebbe in spessore e conducibilità lungo il confine tra nucleo e mantello, ben si accorda con la ipotesi di Buffett.

Il secondo effetto osservabile in superficie è l'influenza esercitata dalla regione nucleo-mantello sul campo magnetico terrestre. L'origine del campo geomagnetico principale è ben compresa, almeno in termini generali (si veda l'articolo *L'evoluzione del campo magnetico terrestre* di Jeremy Bloxham e David Gubbins in «Le Scienze» n. 258, febbraio 1990). È l'effetto dinamo, e non le proprietà magnetiche del ferro presente nel nucleo, a produrre il campo geomagnetico. (Il ferro perde infatti le sue proprietà magnetiche sia alle temperature sia alle pressioni esistenti nel nucleo.) Il rimescolamento del nucleo esterno di metallo liquido agisce essenzialmente come una corrente elettrica che fluisca attraverso un conduttore: analogamente, il nucleo genera un campo magnetico al suo intorno.

La convezione fornisce energia al movimento del nucleo esterno fuso. Il fluido caldo risale verso la sommità del nucleo, con un movimento che trasferisce calore verso l'alto e dà origine a un flusso convettivo. Il materiale fluido più freddo che proviene dalle vicinanze del confine tra nucleo e mantello sprofonda e contribuisce in tal modo anch'esso ad alimentare la convezione del nucleo esterno. È possibile l'esistenza di ulteriori sorgenti di convezione, come la separazione interna tra parte solida e liquida nel nucleo esterno. In questo modo, l'energia meccanica della convezione viene convertita in energia magnetica.

I princìpi che governano questo processo rientrano nella cosiddetta magnetoidrodinamica, combinazione di idrodinamica e di elettromagnetismo. Le equazioni matematiche che regolano questo processo, tuttavia, sono così complesse che nessuno è stato finora in grado di risolverle integralmente. Come risultato, le soluzioni ottenute sono basate su ipotesi plausibili, ma molto semplificate. Le soluzioni ottenute sulla base di queste ipotesi non sono in grado di spiegare i piccoli, ma osservabili dettagli del campo magnetico terrestre, come le tenui increspature nella intensità del campo. Forse la discrepanza risulta da una delle tradizionali semplificazioni utilizzate nei calcoli: quella che considera il nucleo metallico come circondato da una regione elettricamente isolante, corrispondente al mantello. I geofisici ora riconoscono che la parte inferiore del mantello non è completamente isolante, ma consiste di un miscuglio eterogeneo di leghe metalliche e di silicati isolanti.

Incoraggiato da ciò, Friedrich H. Busse della Università di Bayreuth, in Germania, ha riesaminato di recente le equazioni magnetoidrodinamiche. Egli ha scoperto una classe completamente nuova di soluzioni al problema del fenomeno della dinamo terrestre che risultano direttamente dalle variazioni di conducibilità elettrica che si hanno nel mantello inferiore. Le soluzioni dipendono da due fattori principali. Uno consiste nel fatto che le linee del campo geomagnetico sono essenzialmente «congelate» nel metallo liquido del nucleo esterno: così, bloccate al loro posto, le linee del campo si muovono solo con il flusso convettivo del nucleo esterno liquido. Il secondo fattore consiste nel fatto che le regioni metalliche incorporate all'interno dello strato D″ interferiscono con il movimento orizzontale delle linee del campo magnetico che fuoriescono dal nucleo. Lo strato D″ può quindi deflettere o impilare le linee del campo provenienti dal nucleo. Secondo i calcoli di Busse, ambedue i fattori creerebbero campi magnetici locali alla base del mantello, che spiegherebbero parecchie complessità del campo geomagnetico, comprese le increspature nell'intensità del campo.

Le caratteristiche elettromagnetiche del confine nucleo-mantello possono anche influenzare le inversioni del campo magnetico terrestre (si veda l'articolo *Le inversioni del campo magnetico terrestre* di Kenneth A. Hoffman in «Le Scienze» n. 239, luglio 1988). Durante le inversioni, che si producono a intervalli di 100 000 anni, sembra che i poli magnetici seguano una traiettoria preferenziale. Ciò sembra particolarmente evidente per le inversioni più recenti. S. Keith Runcorn dell'Imperial College di Londra e dell'Università dell'Alaska ha postulato parecchi meccanismi sulla base dei quali le variazioni elettriche dello strato D″ potrebbero influenzare il cammino dei poli magnetici.

In un certo senso, quindi, la dinamica tra nucleo e mantello si estende oltre la Terra, allargandosi nello spazio attraverso il campo geomagnetico. Noi ora riconosciamo l'importanza planetaria dell'interfaccia nucleo-mantello, e in futuro diverremo certamente in grado di chiarire come questa remota regione condizioni l'evoluzione del nostro pianeta.

Lo strato D″ si forma per effetto delle reazioni chimiche che avvengono fra nucleo e mantello. Essenzialmente la roccia del mantello fonde parzialmente nel ferro liquido del nucleo esterno producendo sedimenti ricchi di metalli che si depositano al confine fra nucleo e mantello. La convezione del mantello tende a disperdere i prodotti al di sotto delle zone di sprofondamento e ad accumulare materiale in corrispondenza delle risalite. Uno strato sottile arricchito di ossigeno e forse anche di silicio e magnesio potrebbe esistere sul lato interno dell'interfaccia nucleo-mantello.

La cella a incudini di diamante

Questo dispositivo è in grado di riprodurre le pressioni e le temperature esistenti nelle profondità della Terra. Il materiale da comprimere e riscaldare viene posto al centro di una lamina di metallo fra le punte di due incudini di diamante. La rotazione manuale di una vite fa avvicinare le incudini, comprimendo il campione. Un raggio laser può venire focalizzato attraverso il diamante per riscaldare il campione. I profili di composizione mostrano l'abbondanza di ferro, ossigeno, silicio e magnesio (elementi presenti al confine fra nucleo e mantello) prima e dopo il riscaldamento. Le loro quantità sono state riportate in funzione della posizione dell'elemento su una delle superfici del diamante, con la distanza misurata a partire dal bordo. Dopo il riscaldamento, l'area dell'interfaccia si è allargata fino a raggiungere i 10-15 micrometri. La sovrapposizione indica che gli elementi hanno reagito e si è formato una miscela di leghe metalliche (FeSi e FeO) e di ossidi isolanti ($MgSiO_3$ e SiO_2).

BIBLIOGRAFIA

LAY THORNE, *Structure of the Core-Mantle Transition Zone: A Chemical and Thermal Boundary Layer* in «EOS: Transactions, American Geophysical Union», 70, n. 4, 24 gennaio 1989.

JEANLOZ RAYMOND, *The Nature of the Earth's Core* in «Annual Review of Earth and Planetary Sciences», 18, pp. 357-386, 1990.

KNITTLE E. e JEANLOZ R., *Earth's Core-Mantle Boundary: Results of Experiments at High Pressures and Temperatures* in «Science», 251, pp. 1438-1443, 22 marzo 1991.

JACOBS JOHN A., *Deep Interior of the Earth*, Chapman & Hall, 1992.

Le medicine alternative

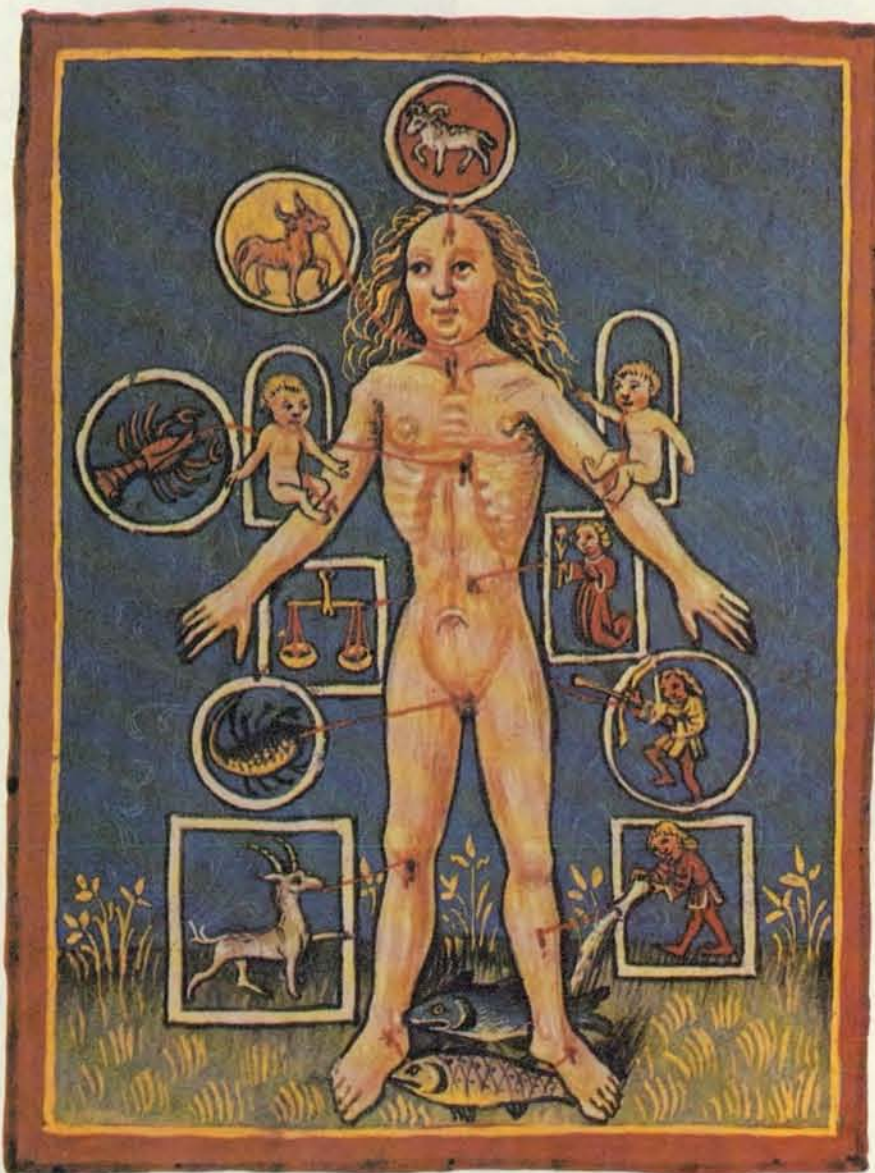
Nonostante la diffusione presso il pubblico, omeopatia, agopuntura e pranoterapia, mancando dei requisiti di sistematicità, oggettività e riproducibilità, non si possono considerare vere discipline scientifiche

di Giovanni Federspil e Cesare Scandellari

Negli ultimi anni si è assistito in Italia e in genere in tutti i paesi occidentali, a una enorme diffusione delle più varie forme di medicina alternativa. Tale fenomeno è stato accompagnato da una vera esplosione dell'interesse collettivo verso le diverse dottrine che sono alla base delle singole pratiche terapeutiche.

Per quanto possa apparire strano che, in un periodo in cui le scienze biomediche stanno mietendo tanti successi, l'attenzione pubblica sia rivolta verso queste pratiche mediche alternative, è tuttavia necessario riconoscere che questa attenzione rientra in quel più vasto movimento di opinione odierno che guarda alla scienza con sospetto o addirittura con ostilità e che, dando credito alle più varie forme di irrazionalismo, vagheggia un ritorno a un immaginario regno della natura incontaminata (si veda l'articolo *Tendenze antiscientifiche nell'Unione Sovietica* di Sergej Kapitza in «Le Scienze» n. 278, ottobre 1991).

Nella «medicina alternativa» sono comprese svariate dottrine e/o numerose pratiche mediche accomunate soltanto dal fatto di essere estranee alla medicina scientifica. Indicate con vari nomi, medicine eretiche, medicine naturali, medicine non-convenzionali, medicine dolci, medicine parallele, medicina non-ufficiale, si fondano su dottrine e prassi che hanno uno statuto scientifico molto differente; mentre infatti alcune di esse sono vere e proprie pratiche empiriche - come l'erboristeria e la chiropratica - , altre affondano le loro radici in un pensiero molto lontano dalla impostazione razionale del pensiero scientifico. Altre medicine eretiche sembrano, infine, confinare con le tradizioni esoteriche o addirittura con la magia.

In questo articolo verranno descritte tre fra le principali medicine eretiche oggi diffuse: l'omeopatia, l'agopuntura cinese e la pranoterapia. Esporremo poi alcune considerazioni intorno al valore di tali pratiche mediche in rapporto alle conoscenze scientifiche che attualmente possediamo.

Secondo l'astrologia medica, giunta attraverso il Medioevo fino a Paracelso, non solo il destino, ma anche la costituzione fisica e le malattie sono regolati dagli astri. (Dal Codice C 54 risalente al 1470, conservato nella Biblioteca Centrale di Zurigo.)

La medicina omeopatica

Fra le medicine alternative, quella meno lontana dalla scienza moderna è, sia per la sua origine sia per il suo contenuto, l'omeopatia.

L'omeopatia nacque in Germania verso la fine del XVIII secolo per opera di un medico sassone, Samuel Hahnemann (1755-1843). Per comprenderne in modo adeguato la natura, è indispensabile conoscere qual era la situazione generale della medicina nel momento in cui questa nuova dottrina vide la luce.

Fra la fine del XVIII e i primi decenni del XIX secolo la medicina europea era andata incontro a una serie di cambiamenti concettuali che ne avevano mutato profondamente l'assetto. Dopo che l'opera di Giovan Battista Morgagni (1682-1771) e quella di Albrecht von Haller (1708-1777) avevano dato vita all'anatomia patologica e alla fisiologia moderna, la clinica basata sugli antichissimi precetti di Galeno mostrava chiaramente tutta la propria insufficienza. In questo nuovo clima i medici cercavano di svincolarsi dalla tradizione e di battere nuove vie che permettessero di rinnovare sia la teoria sia la prassi della medicina. In tutta l'Europa vennero concepite e proposte numerose dottrine, chiamate «sistemi medici», che tendevano a interpretare l'enorme varietà dei fenomeni morbosi sulla base di poche idee molto semplici. Queste teorie, spesso basate su speculazioni filosofiche, risentivano in misura più o meno evidente del clima culturale europeo che, all'epoca, era fortemente influenzato dalla filosofia tedesca della natura. Esse erano caratterizzate da pochi concetti schematici che permettevano di spiegare facilmente un grandissimo numero di fatti clinici e si presentavano come sistemi di idee coerenti e onnicomprensivi che venivano formulati in modo definitivo e che erano quindi del tutto incapaci di venire corretti dall'esperienza. I principali di questi sistemi furono il brownismo, originatosi in Inghilterra, il mesmerismo e il broussarismo in Francia, la dottrina del controstimolo di Giovanni Rasori (1766-1837) in Italia, la frenologia di Franz Joseph Gall (1758-1828) e, appunto, l'omeopatia di Hahnemann in Germania e in Francia.

Nella formulazione originale che ne diede Hahnemann, l'omeopatia si basa su una serie di principi fisiologici, patologici, farmacologici e diagnostico-terapeutici che ne fanno una dottrina unitaria e complessa. Nell'ultima edizione della sua opera maggiore, l'*Organon*, Hahnemann scrive che il funzionamento dell'organismo è dovuto a una «energia vitale immateriale che anima la parte materiale del corpo umano e mantiene l'armonia fra le varie componenti dell'organismo». La malattia in generale, e più specificamente le diverse malattie croniche, sono dovute «a un turbamento di questa energia vitale immateriale», turbamento che deriva dal difettoso stile di vita dell'uomo malato.

A questa concezione generale sulla natura delle malattie Hahnemann affiancò anche una propria nosografia, vale a dire una classificazione e una descrizione delle varie malattie croniche. Queste sarebbero soltante tre: la psora, la lue e la sicosi. La prima fra queste - la psora - è la più grave e la più diffusa delle malattie, quella che provoca tutte le varie affezioni che la medicina identifica e descrive come malattie distinte: dalla cataratta ai disturbi spastici, dalla gotta alla sterilità, dai tumori benigni e maligni alle emottisi.

Nella prassi medica dominante del XVIII e del XIX secolo, in presenza di una certa malattia venivano somministrati quei farmaci che erano in grado di combattere i sintomi del paziente e tale strategia terapeutica veniva sintetizzata nel principio galenico «*contraria contrariis curantur*». Hahnemann, partendo da alcune osservazioni occasionali, ritenne di aver constatato che ogni medicamento, quando veniva somministrato nelle dosi usuali a un soggetto sano, provocava una serie di sintomi che riproducevano quelli osservabili in certe malattie. Partendo da questa constatazione egli propose di somministrare nelle diverse forme morbose proprio quei farmaci che, nel soggetto sano, provocavano la comparsa di una sintomatologia uguale a quella della malattia in esame.

Così, mentre per esempio i medici suoi contemporanei in un paziente febbricitante somministravano un farmaco capace di ridurre la febbre, Hahnemann prescriveva a quel paziente un farmaco capace di provocare nel sano una sintomatologia febbrile. In tal modo egli oppose al comune principio terapeutico dei contrari un principio opposto, il principio dei simili, sintetizzato nell'aforisma «*similia similibus curantur*». Sulla base di tale distinzione, alla comune pratica medica diede il nome di «allopatia» e alla propria dottrina quello di «omeopatia».

Al principio dei simili Hahnemann ne aggiunse presto un secondo, quello delle diluizioni infinitesimali. Sperimentando nell'uomo vari medicamenti, egli ritenne di individuare una nuova legge terapeutica secondo la quale l'effetto di un medicamento, con il ridursi delle dosi, invece di diminuire aumentava progressivamente. Se quindi una sostanza terapeutica veniva diluita sciogliendone un grammo in 10 o in 100 parti di acqua e se poi si diluiva 1 millilitro di questa prima soluzione in 10 o in 100 millilitri di acqua e si ripeteva quest'ultima operazione per 10, 20, 30 o più volte, si potevano raggiungere effetti terapeutici estremamente potenti. Infine, a questo secondo principio Hahnemann ne aggiunse un terzo, quello della dinamizzazione. Egli ritenne, infatti, di avere scoperto che l'effetto terapeutico di un farmaco diluito poteva venire ulteriormente aumentato sottoponendo il farmaco stesso a una serie di scuotimenti manuali.

Al di là di questi principi farmacolo-

Il mesmerismo fu uno dei sistemi medici diffusi in Europa alla fine del XVIII secolo. Esso ebbe molto successo tra le classi agiate, grazie soprattutto alla suggestione esercitata dallo stesso Franz Anton Mesmer, qui ritratto con il suo *baquet*.

Secondo la teoria omeopatica di Samuel Hahnemann le malattie si curano con le sostanze che nel sano provocano la stessa patologia. Questi farmaci omeopatici si trovano presso il Museo di storia della medicina dell'Università di Roma.

gici, Hahnemann modificò profondamente anche la pratica clinica, cosicché ancora oggi il procedimento clinico omeopatico si presenta radicalmente diverso da quello della medicina scientifica. Mentre il medico comune si sforza di formulare una diagnosi quanto più possibile esatta, il medico omeopatico ortodosso non ha alcun bisogno di raggiungere una diagnosi precisa basata sulla comune classificazione delle malattie. La medicina omeopatica, infatti, non attribuisce alla malattia una realtà piena e sottolinea soprattutto l'esistenza degli uomini malati; l'interesse del medico omeopatico è quindi rivolto, più che a diagnosticare una malattia, ad analizzare la sintomatologia del malato e a individuare il farmaco *simillimum*, il farmaco cioè che, provocando nel sano sintomi uguali a quelli del paziente, porterà quest'ultimo alla guarigione.

Si comprende quindi perché la semeiologia del medico omeopatico non abbia nulla a che fare con quella scientifica che è fondata sulla microbiologia, sulla anatomia patologica e sulla fisiopatologia moderna. Per l'omeopata i segni rilevanti non consistono in modificazioni della forma o della struttura degli organi né in alterazioni funzionali o biochimiche, ma nei disturbi del malato che riproducono quelli osservati dopo la somministrazione di questo o quel farmaco. Egli cioè cerca la «patogenesi del rimedio».

Una volta compiuta la «diagnosi individuale», cioè una volta accertato che i disturbi presentati da un paziente sono uguali o almeno molto simili a quelli provocati da un certo medicamento, il medico omeopatico passa alla fase terapeutica somministrando quella sostanza a dosi infinitesimali.

Da questa breve esposizione della dottrina di Hahnemann risulta evidente che nella prassi omeopatica l'attenzione maggiore del medico è rivolta verso la fase terapeutica e che la diagnosi di malattia vi svolge un ruolo secondario. Sul piano storico ciò non deve stupire, poiché costituisce una caratteristica di tutti i sistemi medici del primo Ottocento, nei quali la classificazione delle malattie e la diagnostica corrispondente erano state ipersemplificate, e l'impegno clinico era tutto concentrato nella scelta dei provvedimenti terapeutici.

Sulla base della costruzione teorica che aveva elaborato, il medico sassone continuò a praticare la medicina peregrinando con alterne fortune in varie città tedesche fino a giungere a Parigi, dove ottenne una notevolissima fama professionale e sociale.

Fortemente osteggiato dal mondo della medicina ufficiale, Hahnemann seppe radunare intorno a sé un gruppo di allievi entusiasti i quali iniziarono a diffondere l'omeopatia in tutti i paesi d'Europa. Dopo un primo periodo di notevole successo, nella seconda metà dell'Ottocento essa conobbe una lunga eclissi e soltanto intorno al 1920 ricominciò a diffondersi nella maggior parte dei paesi occidentali.

Dopo la scomparsa di Hahnemann, la diffusione dell'omeopatia si accompagnò alla nascita di accesi contrasti fra gli omeopati, e questi contrasti finirono poi per dare origine a differenti indirizzi teorico-pratici e ad altrettante scuole. Negli Stati Uniti Costantino Hering (1800-1880) fondò una scuola secondo la quale i malati dovevano essere curati con gli stessi prodotti patologici della loro malattia, come, per esempio, il pus gonococcico nella cura della gonorrea. Sempre negli Stati Uniti J. Tyler Kent (1849-1916) accentuò il carattere spiritualistico della dottrina hahnemanniana originale, mentre la scuola francese, con Léon Vannier, sviluppò la dottrina delle costituzioni, identificando tre tipi fondamentali: carbonico, fluorico e fosforico. Fra gli esponenti della scuola argentina, Tomas Pablo Paschero ha sostenuto che la malattia è una perturbazione dello spirito originata da uno squilibrio fra intelletto e volontà, mentre Alfonso E. Masi ha identificato la psora con la vulnerabilità e l'irritabilità delle cellule.

Nel 1952 Hans Heinrich Reckeweg, cercando di avvicinare l'omeopatia ai concetti della biochimica moderna, ha proposto una nuova dottrina, denominata omotossicologia, basata sull'assunto che tutte le aggressioni provenienti dall'ambiente sono dovute all'azione di non meglio precisate omotossine.

Sul piano pratico, poi, gli omeopati si sono divisi in unicisti, in pluralisti e in complessisti, a seconda che ad un paziente venga prescritto un solo farmaco o più farmaci in sequenza o, infine, più farmaci contemporaneamente.

Le varie scuole omeopatiche oggi esistenti si differenziano fra loro per il diverso atteggiamento che assumono nei riguardi della dottrina originaria. Alcuni infatti affermano di volere una omeopatia che tenga conto dei progressi delle varie scienze biomediche e che collabori con la medicina ufficiale, mentre altri non intendono in alcun modo di doversi allontanare dalla teoria e dalla prassi elaborate da Hahnemann. Tuttavia, al di là dei contrasti che oppongono una scuola all'altra e al di là delle proclamazioni verbali di adesione alla scienza moderna, tutti i medici omeopatici si riconoscono nel ritenere validi i tre principi fondamentali enunciati da Hahnemann, la legge dei simili, il principio delle diluizioni infinitesimali e la dinamizzazione del rimedio omeopatico.

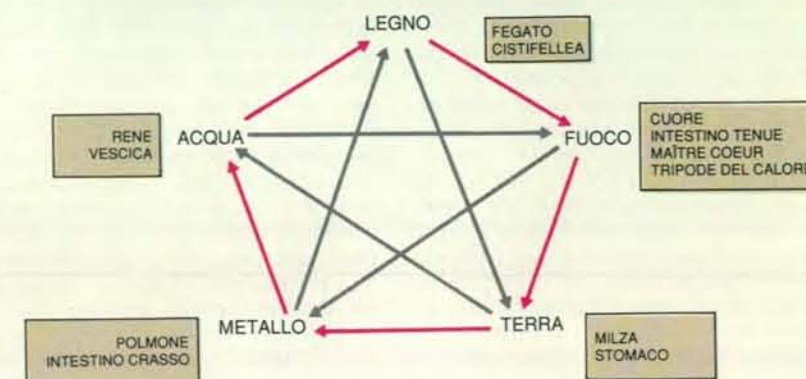
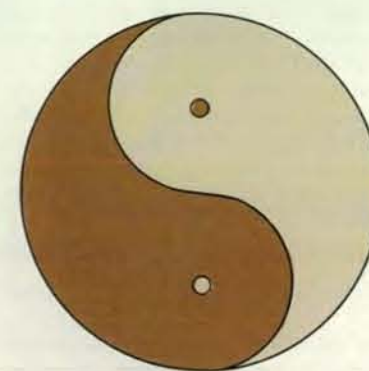
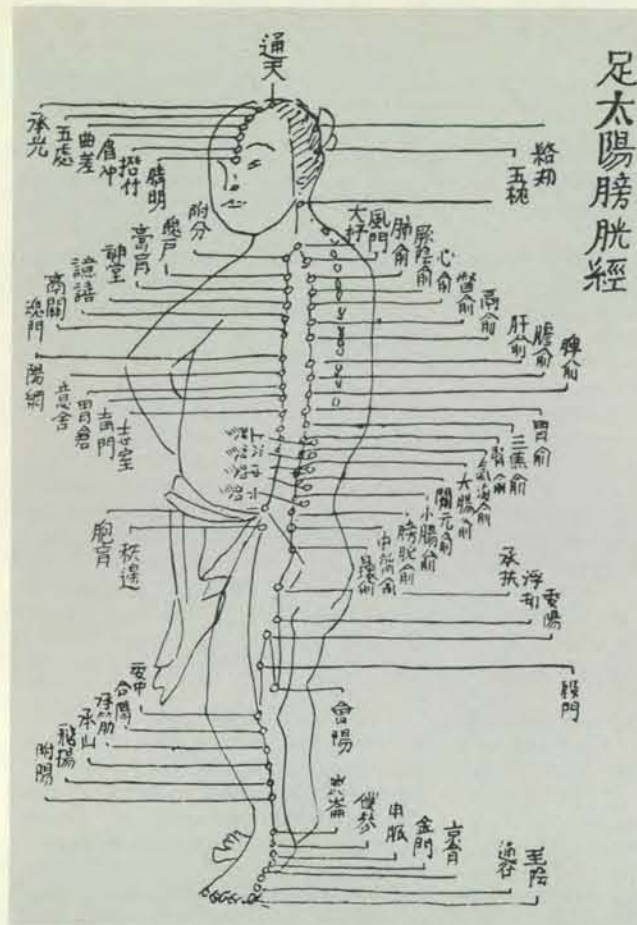
Una analisi approfondita delle varie tesi dell'omeopatia richiederebbe un discorso molto lungo, tuttavia anche questa rapida esposizione permette di svolgere alcune considerazioni critiche.

Nello sviluppo del pensiero medico la separazione fra sapere scientifico e sapere extra-scientifico è avvenuta lentamente e con grande difficoltà lungo tutto l'arco del XVIII e del XIX secolo. A cavallo fra questi due secoli, i sistemi medici, che influenzavano buona parte del mondo accademico e di quello professionale, rappresentavano dottrine e pratiche mediche in cui le affermazioni scientifiche e quelle filosofiche erano ancora intrecciate molto strettamente. L'omeopatia, sorta come uno dei tanti sistemi, si è sempre mantenuta fedele ai principi teorici originari e si presenta ancora oggi come una teoria medica nella quale tesi scientifiche e tesi metafisiche sono strettamente commiste.

Analizzandone il contenuto è facile constatare che già sul piano della chimica le idee dell'omeopatia sono del tutto insostenibili. Quando infatti si diluisce una sostanza con la tecnica omeopatica si ottengono rapidamente diluizioni così elevate che, dopo pochi passaggi, non rimane che solvente puro. Come è ben noto, grazie al numero di Avogadro (la costante che indica il numero di molecole presenti in una grammomolecola di qualsiasi sostanza), conoscendo i grammi di una sostanza sciolti in una soluzione e il peso molecolare della sostanza stessa è possibile calcolare il numero delle molecole presenti in quella soluzione. Un farmacologo italiano, Luigi Sabbatani, ha dimostrato nel 1921 che, a partire dalla tredicesima diluizione centesimale, nelle preparazioni omeopatiche non esiste che solvente puro. Se si considera, quindi, che una delle diluizioni più usate è la trentesima e che gli omeopati possono giungere fino alla duecentesima, si può ben comprendere quanto sia assurdo ritenere che in questi preparati con l'aumentare della diluizione aumenti l'azione del rimedio.

Di fronte a questa confutazione della loro teoria gli omeopati, per spiegare i pretesi effetti terapeutici dei loro rimedi, hanno fatto ricorso a una serie di ipotesi fantasiose, prive di ogni supporto sperimentale. Mentre alcuni hanno sostenuto che la «dinamizzazione» del farmaco ottenuta mediante una serie di scosse libererebbe l'energia contenuta nell'atomo, altri ritengono che l'attività delle alte diluizioni sarebbe dovuta a un riaggiustamento molecolare del solvente, riaggiustamento che sarebbe specifico del principio attivo e del livello di diluizione considerato.

Il dibattito suscitato pochi anni or sono dai risultati ottenuti dal francese Jacques Benveniste sulla cosiddetta «memoria dell'acqua» ha avuto un'eco così larga perché da più parti quei risultati sono stati interpretati come una prova a favore dell'omeopatia. Senza entrare nella polemica sulla correttezza degli esperimenti, è evidente che quella ricerca, anche se avesse provato che l'acqua possiede una «memoria», dimostrerebbe solo la possibilità che le diluizioni infinitesimali agiscono e non costituirebbe in alcun modo la conferma di una dottrina complessa come l'omeopatia.

Anche sul piano più strettamente biomedico la credibilità dell'omeopatia non è migliore. Essa non solo adopera i farmaci a diluizioni prive di significato, ma non usa solitamente i medicamenti che

Nella filosofia cinese l'esistente è governato da due forze opposte e interagenti, yin e yang, rappresentate dal famoso simbolo riprodotto qui sopra. Lo schema a destra evidenzia i due cicli creativo o cheng (*in rosso*) e distruttivo o k'eu (*in grigio*) in base ai quali gli elementi e gli organi, attraverso il flusso di energia, si influenzano favorevolmente o sfavorevolmente.

Le illustrazioni, tratte da un antico manuale cinese di agopuntura, mostrano i meridiani della vescica (*a sinistra*) e dei reni (*a destra*). L'infissione di aghi ha lo scopo di influire sull'energia circolante nei meridiani e quindi negli organi collegati.

la farmacologia moderna ha messo a disposizione della clinica. Restando ancorata agli usi e alla terminologia della prima metà dell'Ottocento, essa impiega come farmaci solo le sostanze minerali e vegetali che venivano adoperate al tempo di Hahnemann. La cosiddetta ricerca farmacologica e la patogenesi dell'omeopatia non hanno poi nulla in comune con la farmacologia e la fisiopatologia della medicina scientifica. L'azione di un rimedio non viene infatti descritta attraverso i suoi meccanismi fisiologici o biochimici, ma solo attraverso i segni che quel rimedio provoca nel soggetto sano. Considerando poi i sintomi attribuiti dagli omeopati a ciascun farmaco, si può constatare che il loro numero è enorme e che non hanno alcuna relazione con quanto è abitualmente descritto nella letteratura scientifica. Per esempio, nella prima edizione dell'*Organon*, al rimedio denominato *Pulsatilla nigricans* venivano attribuiti 309 sintomi, mentre, dopo soli 17 anni, nella seconda edizione, i sintomi erano passati a 1164 e nel volume di un altro autore essi erano addirittura 1323. Dai sintomi provocati da un farmaco gli omeopati, poi, ritengono di poter inferire l'esistenza di tipi umani a esso sensibili e di tipi corrispondenti alle azioni del farmaco stesso. Così, per esempio, un medico omeopata dei nostri giorni ha scritto che «il soggetto *Pulsatilla* si caratterizza per una atonia congenita o acquisita delle pareti venose, specie dei capillari» e che «in *Pulsatilla* tutto ciò che aumenta la congestione venosa peggiora il suo stato mentale». Quindi «le sue idee, così come il suo umore, sono capricciose, mutevoli e instabili... *Pulsatilla* è dolce, tenera, servizievole, timida e romantica... *Pulsatilla*, che è proprio l'opposto del tipo collerico, può anche divenire aggressiva, acida, violenta ed esplodere in crisi di rabbia... L'ipocrinia, specie ovarica, orienta sensibilmente il suo atteggiamento sessuale... Se si sposa diviene gelosa proprio perché non ha fiducia nelle sue limitate capacità sessuali».

In questo modo si è venuta elaborando una dottrina costituzionalistica tanto fantastica quanto arbitraria, che i medici omeopatici tengono in gran conto nella loro pratica quotidiana.

Nell'ambito della patologia, la dottrina omeopatica ignora totalmente le acquisizioni conseguite dalla scienza negli ultimi 200 anni. Nei manuali omeopatici, infatti, si cercherebbero invano riferimenti precisi all'anatomia e all'istologia patologica, alla microbiologia o alle alterazioni endocrino-metaboliche che rappresentano tanta parte del sapere medico contemporaneo.

Quando poi, come è avvenuto in qualche occasione, si è cercato di introdurre le idee omeopatiche nell'ambito delle conoscenze fisiopatologiche moderne, si è ottenuto soltanto un miscuglio confuso di concetti e di affermazioni del tutto privi di qualsiasi coerenza interna.

Passando infine alla terapia, ancora oggi non è stata fornita alcuna giustificazione razionale del principio hahnemanniano dei simili. In realtà, mentre il principio dei contrari, non avendo più alcun senso nell'ambito della medicina moderna, è stato abbandonato, i medici omeopatici continuano a considerare la legge dei simili come il cardine di ogni terapia. Tale legge, tuttavia, non trova alcuna giustificazione nelle conoscenze fisiopatologiche attuali e può al massimo costituire una regola dettata dall'esperienza terapeutica.

Allo scopo di sostenere la validità del principio «*similia similibus curantur*», i seguaci dell'omeopatia si sono appellati alla tecnica delle vaccinazioni, sostenendo che, con il suo impiego, la medicina scientifica non fa altro che applicare la dottrina omeopatica. Anche se nelle apparenze l'analogia può sembrare stretta, non altrettanto si può dire della sostanza. Le vaccinazioni hanno l'intento non tanto di riprodurre i sintomi della malattia (come vuole la teoria omeopatica), quanto quello di indurre nell'organismo vaccinato la produzione di anticorpi capaci di contrastare l'agente patogeno. E il fatto che l'obiettivo finale della vaccinazione sia l'apporto di anticorpi all'organismo malato è confermato dalla considerazione che, nella pratica terapeutica, la vaccinazione - in realtà utilizzata più per la prevenzione che per la cura delle malattie - è spesso sostituita dalla immunizzazione passiva, che consiste nella somministrazione diretta di anticorpi specifici.

Tra le applicazioni terapeutiche dell'infissione di aghi i maggiori successi riguardano il campo dell'analgesia. Tali risultati non possono però venire considerati una conferma delle teorie su cui la pratica è basata. I volti sorridenti del disegno qui riprodotto sono una chiara promozione dell'agopuntura nell'anestesia chirurgica.

Tuttavia, anche se si volesse ammettere che il principio delle vaccinazioni coincide col principio omeopatico, risulta chiaro che ciò non dimostra la validità del principio di Hahnemann. Mentre infatti quest'ultimo si presenta come un principio terapeutico generale, applicabile a tutte le forme morbose, la vaccinazione rappresenta un criterio profilattico-terapeutico di uso ben circoscritto.

Ma proprio perché il principio dei simili non può essere niente di più di una regola empirica, esso dovrebbe basarsi su prove ineccepibili. Se invece si analizzano le basi cliniche consensuali su cui tale principio viene affermato, si può constatare che, dal tempo di Hahnemann, ben poco cammino è stato compiuto sul piano dell'acquisizione dei fatti. Come le affermazioni terapeutiche riportate nell'*Organon* appaiono del tutto insufficienti per un medico contemporaneo, così la gran parte dei successi terapeutici vantati dai medici omeopatici si rivela scientificamente inconsistente.

L'agopuntura e la medicina cinese

Altrettanto diffusa dell'omeopatia è l'agopuntura cinese. Questa è una dottrina medica, sorta in Estremo Oriente un millennio circa prima di Cristo, e una tecnica terapeutica che, come scrive Joseph Needham in *La medicina cinese* (Il Saggiatore, Milano, 1982), «consiste nella infissione di aghi molto sottili in differenti parti del corpo e in punti precisamente specificati secondo uno schema che ha la struttura di una mappa».

Per quanto l'agopuntura sia spesso presentata isolatamente, essa in realtà fa parte di un gruppo di tecniche terapeutiche apparentate fra loro e basate su principi teorici derivati dall'antica filosofia cinese. Fra queste le più note sono l'auricoloterapia, l'auricolodiagnostica, lo shiatsu e la moxibustione.

Originatasi da osservazioni popolari casuali, l'agopuntura cinese si è andata arricchendo nei secoli di contenuti teorici sempre più complessi. In seguito, venendo a contatto con culture diverse, la teoria e la pratica dell'agopuntura hanno dato vita a molte varianti.

Secondo la dottrina classica che deriva direttamente dal taoismo, l'intero universo è attraversato da una energia primordiale, detta qi o tch'i, che lo anima. Questa energia possiede due aspetti, ovvero due componenti fondamentali, fra loro strettamente interdipendenti, al punto che l'una non può esistere senza l'altra: lo yang e lo yin. Lo yang è ciò che sta sopra rispetto a ciò che sta sotto, ciò che è non-materiale, ciò che è più caldo, più secco, più attivo, più vivo, ciò che sta a destra; lo yin è invece ciò che sta sotto, che è materiale, che è meno caldo, meno attivo, che sta a sinistra e così via.

In realtà, yang e yin non solo sono fra loro contrapposti, ma si compenetrano, in modo tale che nessuno di questi due

La pranoterapia si basa sull'ipotesi che taluni emettano una forte carica di energia bioradiante. Gli esperimenti condotti da Semion D. Kirlian fotografando organismi viventi in presenza di un campo elettrico furono presi a riprova di questa ipotesi, ma in realtà il fenomeno ha una spiegazione fisica. Questa foglia mostra per effetto Kirlian un alone più o meno luminoso a seconda delle stato di salute dei suoi tessuti.

principi è mai assolutamente puro e ognuno dei due contiene al suo interno tracce dell'altro; per questo motivo ogni cosa del mondo è yang ed è contemporaneamente yin: così, per esemplificare, la cintola è yang in rapporto ai piedi ed è yin in rapporto alla testa, e una temperatura di 10 gradi Celsius è yin in rapporto a 30 gradi ed è yang in rapporto a 0 gradi.

Nell'uomo l'energia universale costituisce la «forza vitale» o «pneuma» e circola in continuazione fluendo in canali sottilissimi situati sotto la pelle, i quali per definizione sono inosservabili. Il sistema reticolare di canali (detti meridiani o tratti) in cui la forza vitale (qi, di natura yang) scorre e viene trasportata a tutti i tessuti dell'organismo è costituito dai 12 canali principali (zhengjing), dai canali di intersezione (jingbie), dai canali di congiunzione (luo), dalle diramazioni (luomai, sunluo ed hengluo) «di cui nessuno conosce veramente la disposizione e che devono essere almeno 300». I canali emergono alla superficie del corpo in particolari zone, i punti, che sono sfruttate dai medici agopuntori per modificare il flusso dell'energia. Circa il numero dei punti esistenti regna un rilevante disaccordo: mentre un manuale classico, il *Nei Jing*, in correlazione simbolica con i giorni dell'anno, indicava l'esistenza di 365 punti, i punti erano già diventati 670 nel 1974 e negli ultimi anni il loro numero si è accresciuto.

Accanto all'energia yang che scorre nei canali (sistema jingjin e sistema jingluo), esiste un'energia yin che circola attraverso i vasi sanguigni (sistema jingmai) e che è ben separata dalla prima.

Secondo la medicina cinese, la vita è costituita dalla integrazione dello yang e dello yin e l'equilibrio di queste due forze rappresenta la condizione fondamentale della salute; il loro squilibrio è invece la causa prima della malattia e la loro separazione costituisce la morte. Nell'organismo, quando il flusso dell'energia si svolge regolarmente, si mantiene lo stato di salute; quando invece si verifica uno squilibrio nella circolazione dell'energia vengono prodotte le varie malattie. Se l'equilibrio fra yang e yin non viene mantenuto, si produce una serie di fenomeni, detti «di vittoria» o «di sconfitta», i quali non sono altro che manifestazioni patologiche; così, per esempio, lo yang vittorioso provoca i segni del caldo (febbre) mentre lo yin vittorioso provoca i segni del freddo (brividi).

Gli squilibri energetici possono manifestarsi a carico di ogni organo e possono consistere in un difetto o in un eccesso. Tali squilibri, che si formano nei visceri, vengono trasmessi dai tratti viscerali ai tratti superficiali cutanei dei meridiani. Se si considera che i punti regolano il flusso dell'energia lungo i vari meridiani, si capisce perché nella medicina cinese la stimolazione dei punti mediante aghi rappresenti un momento terapeutico fondamentale. Il compito del medico è solo quello di ripristinare il normale flusso dell'energia nell'organismo e per raggiungere questo scopo egli deve infiggere aghi metallici sottilissimi là dove l'energia scorre, favorendone od ostacolandone il flusso. Per correggere le alterazioni energetiche degli organi è tuttavia necessario conoscere i rapporti di gerarchia o di parentela che collegano un organo all'altro.

L'energia fluisce nell'organismo secondo il seguente ordine ciclico: polmone → intestino crasso → stomaco → milza → cuore → intestino tenue → vescica → rene → maître coeur → tripode del calore → vescichetta biliare → fegato → polmone e così di seguito. In tal modo ogni organo è «figlio» e «nipote» di alcuni visceri e «genitore» e «nonno» di altri. Queste parentele rappresentano un aspetto delle relazioni che esistono fra i cinque elementi primordiali della natura, il fuoco, la terra, il metallo, l'acqua e il legno, e che costituiscono il «ciclo creativo o ciclo Cheng». Accanto a queste esistono poi relazioni inibitorie o di sottomissione fra i vari elementi che costituiscono il «ciclo distruttivo o ciclo k'eu». Nella medicina cinese esiste una corrispondenza fra i vari elementi primordiali e gli organi, cosicché, per esempio, al legno corrispondono il fegato e la cistifellea. Riunendo ora in un solo schema i due cicli e le varie corrispondenze fra elementi e organi, si ottiene l'illustrazione di pagina 35 a destra. Variando l'energia di un organo si potranno quindi influenzare, positivamente o negativamente, gli organi a esso connessi; così, un aumento energetico del fegato produrrà un analogo aumento sul cuore (organo «figlio») e sulla milza e sullo stomaco (organi «nipoti»), mentre un vuoto energetico su un viscere determinerà un'aggressione su quel viscere da parte dell'organo «nonno» e del «genitore». L'agopuntura viene eseguita per «disperdere» (diminuire) o «tonificare» (aumentare) l'energia circolante sul meridiano corrispondente e di conseguenza sul rispettivo organo e sugli organi a esso collegati.

Tralasciando la parte diagnostica della medicina cinese, che è basata sull'esame dei polsi e che è particolarmente complessa, si può dire che gli agopuntori utilizzano in pratica mappe del corpo umano in cui si riconoscono i vari punti. Il medico deve dapprima diagnosticare il tipo di alterazione energetica e la sua localizzazione; da questi elementi deve poi risalire ai meridiani che corrispondono agli organi sofferenti e riconoscere i punti più adatti per apportare l'azione terapeutica; infine deve stabilire se si debba tonificare o disperdere l'energia. Nella pratica vengono impiantati in vari punti fino a 14 aghi d'oro, d'argento o d'acciaio inossidabile, che vengono lasciati *in situ* per 10-20 minuti.

La tecnica dell'infissione degli aghi fu chiamata agopuntura dai gesuiti, che nel XVII e XVIII secolo si erano stabiliti alla corte imperiale cinese. Essa fu fatta conoscere in Occidente da due diplomatici, Dabry de Thiersant (1863) e, soprattutto, G. Souliè de Morant, il quale verso il 1930 suscitò in Europa un grande interesse intorno alla medicina cinese.

Attualmente l'agopuntura, da sola o associata ad altre tecniche terapeutiche di origine orientale, è largamente praticata nei paesi occidentali, dove sono sorte numerose scuole di agopuntura che si contendono il campo. Le indicazioni della terapia con agopuntura cinese, proposte dai vari seguaci di questa dottrina, sono, in linea di principio, vastissime, pur variando notevolmente a seconda degli indirizzi dei diversi terapisti. Secondo gli agopuntori più vicini al pensiero tradizionale cinese l'agopuntura può dare benefici in patologie che vanno dalla tachicardia e dalla cefalea alla febbre, all'emottisi, alla diarrea, all'asma, ai disturbi ipofisari e così via.

Negli autori più vicini al pensiero scientifico si può notare una certa limitazione nelle indicazioni dell'agopuntura che col tempo sono sempre più circoscritte agli effetti analgesici. Secondo quanto viene affermato da più parti, l'azione analgesica della infissione di aghi sarebbe così marcata da consentire di effettuare diversi tipi di interventi chirurgici somministrando basse dosi di farmaci o addirittura senza anestesia.

Tali risultati hanno spinto molti autori a tentare di spiegare l'analgesia da agopuntura sulla base delle attuali conoscenze fisiologiche. Questi studi vengono oggi condotti applicando per lo più concetti e tecniche che non appartengono alla tradizione culturale cinese e che sono invece derivati dalle attuali conoscenze scientifiche. È quindi ovvio che i risultati di queste ricerche - al di là dei nomi che vengono impiegati - non fanno parte dell'agopuntura cinese né possono in alcun modo costituire una convalida sperimentale di questa dottrina.

L'agopuntura cinese si presenta, sul piano concettuale, molto più lontana dell'omeopatia dalla scienza moderna. Essa infatti non solo deriva da una filosofia estremamente distante dal pensiero razionale occidentale, ma fa costantemente uso di idee che non hanno alcun elemento in comune con le conoscenze biomediche attuali. A questo si deve aggiungere che i concetti più importanti a cui essa fa riferimento - l'energia, i canali, i punti - sono per definizione inosservabili e non possono quindi entrare a far parte del discorso scientifico.

Sul piano operativo i medici agopuntori vantano notevoli successi nei campi più disparati della patologia. Nella realtà, gli effetti dell'agopuntura cinese, quando sono stati studiati con i metodi rigorosi della sperimentazione clinica controllata, o si sono rivelati estremamente modesti o addirittura non sono stati confermati. Nelle sperimentazioni più moderne sotto il nome di agopuntura vengono solitamente, ma indebitamente, comprese varie tecniche stimolatorie che vanno dall'agopuntura cinese classica fino alla elettroagopuntura e alla stimolazione nervosa transcutanea. Questo stato di cose ha portato ad avvicinare l'agopuntura tradizionale alla «analgesia elettrica locale», creando confusioni e ambiguità nei termini. L'effetto analgesico ottenuto con varie tecniche agopunturali non è obbligatoriamente legato all'agopuntura cinese e può derivare da modificazioni del metabolismo della serotonina o dei peptidi oppioidi cerebrali. Esso è comunque tuttora in fase di studio.

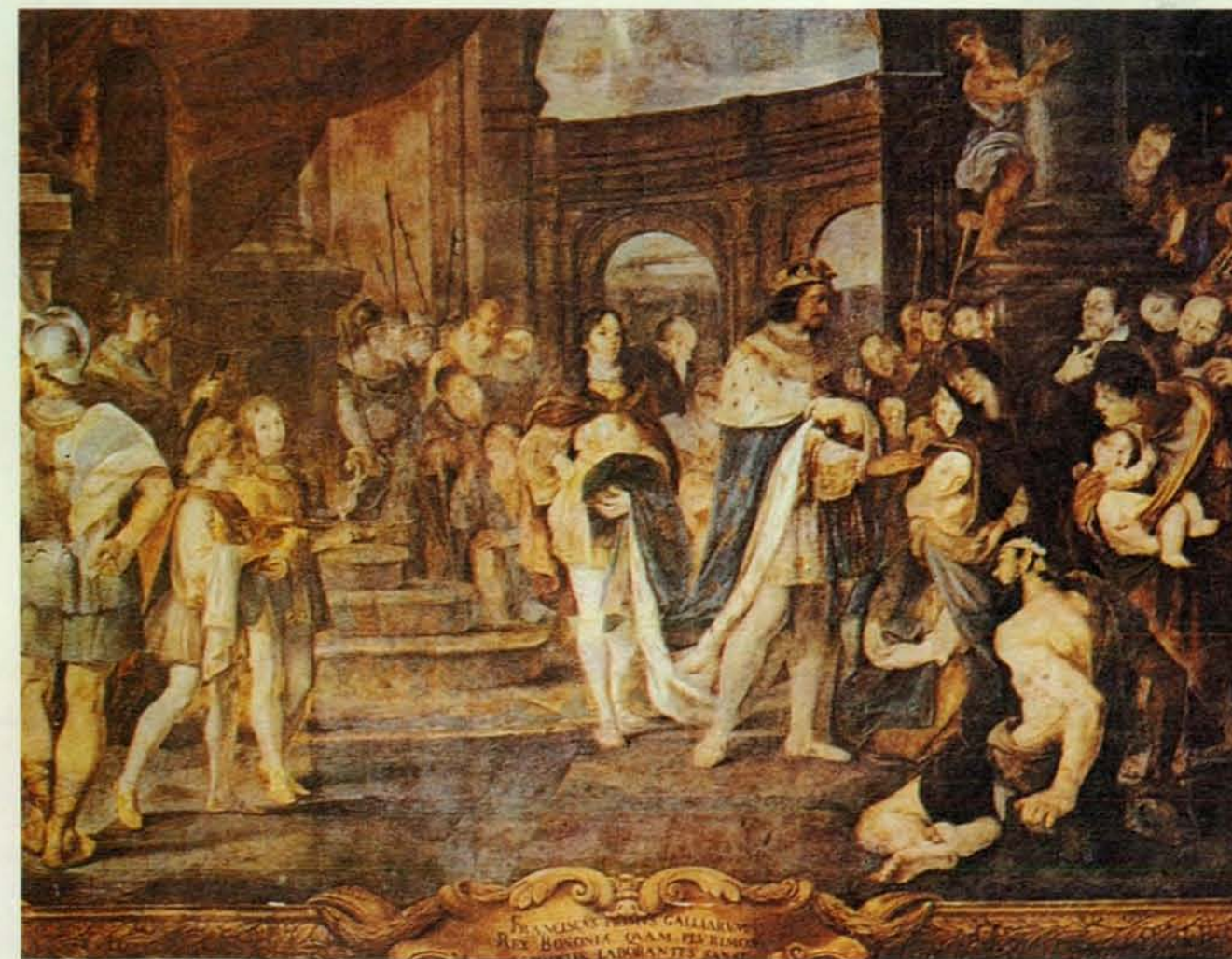
La pranoterapia

Con questo termine si designa l'attività terapeutica svolta da alcuni individui - comunemente chiamati guaritori - che «intendono e ritengono di influire sullo stato di malattia mediante la semplice imposizione delle mani sul corpo del paziente (per contatto o a breve distanza) senza l'uso di altri mezzi».

Sul piano storico, pratiche pranoterapiche sono rintracciabili sia nelle civiltà più antiche, come quella egiziana, sia presso popolazioni molto primitive. Attualmente tale prassi, che fino a pochi decenni or sono era diffusa soprattutto negli ambienti rurali, si è estesa negli ambienti metropolitani. In realtà, il mondo della pranoterapia è poco strutturato sia sul piano sociale sia su quello concettuale, cosicché, per esempio, non esistono associazioni con le caratteristiche delle società scientifiche che riuniscano tutti coloro che attuano tale prassi terapeutica, né esistono pubblicazioni che riportino una descrizione obiettiva e sistematica dei risultati ottenuti dai guaritori.

Gran parte delle pratiche pranoterapiche viene effettuata in ambienti fortemente impregnati di magia o di spiritismo e da parte di persone del tutto prive di una qualsiasi cultura scientifica; tuttavia in alcuni casi esse vengono effettuate da persone in possesso di un regolare titolo di studio in medicina. Di fronte alla quasi totale mancanza di una seria documentazione scientifica, sta invece l'enorme numero di successi che i pranoterapisti vantano nei più diversi campi della patologia. Mentre alcuni fra loro hanno sostenuto di aver guarito soggetti affetti da neoplasie o da cerebropatie organiche o da malattie demielinizzanti o da artrite reumatoide, altri ammettono di ottenere successi solo nelle forme morbose che non sono caratterizzate da lesioni anatomiche evidenti, come le cefalee o l'epilessia o le nevralgie.

Sul piano teorico la spiegazione più antica e più diffusa circa le guarigioni ottenute con l'imposizione delle mani attribuisce il fenomeno al passaggio dalle mani del guaritore al paziente di un *quid* che modificherebbe lo stato di malattia dell'organismo. Tale *quid* è stato variamente identificato con un fluido o con un'energia fisica di natura sconosciuta denominata energia bioradiante. In realtà, il termine pranoterapia deriva dal sanscrito *prana* che significa respiro, vitalità, forza e che non ha nulla a che vedere con una forma di energia.

Finora l'idea del passaggio di un *quid* tra guaritore e malato ha ricevuto differenti versioni. Secondo un noto medico-guaritore, la base della terapia biora-

Il dipinto, che decora la sala Farnese di palazzo d'Accursio a Bologna, raffigura il re di Francia che risana gli scrofolosi. La credenza nel potere taumaturgico di certuni connette la pranoterapia alle pratiche magiche o almeno alla suggestione.

diante sarebbe «il trasferimento di energia biofisica, biopsichica e biospirituale fra medico e malato». Il medico emetterebbe quindi onde bioradianti dotate di energia terapeutica che influenzerebbero sia l'organismo sia la parte psichica e spirituale dell'uomo, portandolo a uno stato di letizia interiore.

Secondo gli autori della scuola russa, attorno a ogni organismo vivente esisterebbe uno stato particolare della materia, detto bioplasma, costituito da un insieme di particelle elementari e privo di struttura atomica. Il bioplasma permetterebbe appunto di trasferire l'energia da un organismo all'altro e spiegherebbe dunque la possibilità delle guarigioni indotte dai pranoterapeuti.

Negli anni settanta le immagini fotografiche di Semion D. Kirlian, che venivano ottenute facendo attraversare la pellicola da una scarica elettrica ad alta tensione e che mostravano l'esistenza di un alone luminoso intorno agli esseri viventi, furono considerate una prova dell'esistenza di una «bio-aura», ovvero di «una specie di involucro pulsante pieno di energia» o anche di «un corpo di bioplasma con caratteristiche semimateriali, che collega il corpo immateriale a quello materiale». In realtà, più tardi si vide che tali immagini si potevano ottenere anche fotografando oggetti inanimati e che erano prodotte dalla ionizzazione dei gas che li circondavano.

Altri studiosi hanno invece attribuito le guarigioni ottenute con l'imposizione delle mani alla psicocinesi, cioè alla capacità della mente di agire sulla materia. Secondo Piero Cassoli e Giovanni Iannuzzo l'effetto placebo e la possibilità di un'azione psicosomatica, pur riuscendo a spiegare un numero notevole di guarigioni pranoterapeutiche, non potrebbero venire invocati in tutti i casi. Fondandosi sui risultati delle ricerche parapsicologiche, essi ammettono che alcuni soggetti possano influenzare un sistema fisico tramite un'azione mentale, ma ritengono che al processo di guarigione possa contribuire anche quella vaga entità che i parapsicologi chiamano «psi».

Come si vede, la situazione della pranoterapia appare del tutto singolare sul piano scientifico. Da un lato infatti si assume che nelle guarigioni ottenute con l'imposizione delle mani intervengano fattori misteriosi e del tutto ignoti alla scienza moderna, dall'altro la base osservativa registrata e descritta secondo i più comuni canoni dell'osservazione scientifica è pressoché assente.

Nonostante siano sorte negli ultimi anni alcune società che intendono promuovere studi obiettivi in questo campo e nonostante l'apertura di credito che alcuni studiosi di varia formazione le hanno dato, la pranoterapia resta estremamente lontana da una qualsiasi ammissibilità scientifica e, per molti aspetti, si accosta alle pratiche magiche.

Fra i pochissimi studiosi che hanno cercato di fornire una documentazione adeguata dei risultati della pranoterapia, Cassoli ha classificato i suoi pazienti in nove classi decrescenti di attendibilità diagnostica e ne ha dato una obiettiva per quanto sommaria descrizione clinica. Analizzando questa casistica con i criteri comunemente adottati in clinica medica, si può constatare che, nella grande maggioranza dei casi, queste osservazioni cliniche non sono in alcun modo dimostrative dell'efficacia della pranoterapia, e che nei casi restanti il miglioramento clinico o non era dimostrato o poteva essere spontaneo.

Naturalmente, va sottolineato che anche se i risultati riportati suggerissero realmente un effetto della pranoterapia, essi non sarebbero ancora dimostrativi sul piano scientifico. È infatti evidente che i casi riportati sono stati scelti da una casistica molto più ampia ed è risaputo che una valutazione scientificamente rigorosa dei risultati di una qualsiasi terapia impone di presentare e di analizzare tutti i risultati ottenuti e non solo quelli che hanno avuto esito favorevole.

Come valutare le medicine alternative?

Come si è visto, le medicine alternative considerate fin qui sono tutte costituite da teorie che non possiedono punti di contatto con la scienza contemporanea o addirittura si presentano in aperto contrasto con quelle che al momento attuale sono le conoscenze scientifiche più consolidate e accreditate.

La prima caratteristica delle medicine alternative che colpisce l'osservatore è la loro aperta violazione del principio di sistematicità. Una caratteristica fondamentale della conoscenza scientifica è infatti la sua tendenza alla unitarietà, alla elaborazione cioè di un sapere sistematico che colleghi le varie discipline in un corpo integrato e coerente di conoscenze. Nel mondo delle medicine eretiche tale caratteristica è del tutto assente: ogni singola medicina, infatti, vive di una vita propria e non si cura di stabilire connessioni stabili fra le proprie teorie e quelle della medicina scientifica o quelle delle altre medicine non-scientifiche.

Analizzando ogni singola dottrina si può notare una tendenza che è esattamente opposta a quella della scienza autentica. Mentre infatti in questa i ricercatori tendono a ridurre progressivamente il dissenso, creando teorie sempre più ampie che tengano conto di tutti i fatti che si accumulano, nel mondo delle medicine alternative ogni dottrina si occupa esclusivamente dei fenomeni che concernono il proprio ambito e ignora totalmente i fenomeni che riguardano le altre dottrine. In tal modo si vanno creando sempre nuovi indirizzi e nuove scuole che, isolandosi e differenziandosi sempre più, fanno aumentare progressivamente il dissenso fra i loro aderenti.

Questo comportamento dei seguaci delle medicine alternative viola anche un'altra caratteristica basilare della conoscenza scientifica: la ricerca della consensualità e della oggettività. Come ha ben sottolineato il fisico John Ziman, «la oggettività della conoscenza scientifica risiede nel fatto che essa è un costrutto sociale che non deve la sua origine a qualche particolare individuo, ma che risulta invece da una creazione cooperativa e comunitaria».

Al di là delle questioni riguardanti il loro contenuto, le medicine alternative sono caratterizzate da numerosi vizi metodologici che le rendono radicalmente diverse dalle vere discipline scientifiche.

Prima di tutto le osservazioni fattuali, vale a dire i resoconti osservativi su cui esse si basano, sono quasi sempre caratterizzate da una estrema genericità e da una assoluta imprecisione. A volte, come avviene nel caso dell'agopuntura cinese, la massima parte dei dati sugli effetti terapeutici viene fatta risalire a osservazioni compiute molti secoli or sono, e si assume che la lunghissima tradizione popolare equivalga all'osservazione scientifica e sia, di per se stessa, garanzia di veridicità.

In molti casi i successi vantati dagli aderenti alle varie medicine alternative concernono osservazioni aneddotiche, nelle quali l'esistenza di un rapporto di causalità fra provvedimento terapeutico e variazioni del quadro clinico è almeno fortemente opinabile. Quando poi gli effetti terapeutici riguardano gruppi di soggetti, spesso non viene fornita alcuna descrizione precisa sia della casistica studiata sia dei fenomeni osservati. Infine, nelle pubblicazioni delle medicine eretiche l'analisi statistica dei risultati o è completamente assente o è del tutto insufficiente e/o scorretta.

Se si considerano le teorie delle medicine alternative è facile constatare che esse non possiedono in alcun modo il carattere delle vere teorie scientifiche. Ciò appare subito evidente dal fatto che i concetti su cui si basano non sono in alcun modo definibili operativamente; l'energia primordiale, lo yang, lo yin, il punto dell'agopuntura cinese, l'energia vitale immateriale, la dinamizzazione, la psora, i miasmi della medicina hahnemanniana, il fluido, il prana, il corpo astrale, le emissioni bioelettriche della pranoterapia sono altrettanti esempi di concetti e di fenomeni che, nella forma in cui sono stati e vengono ancor oggi formulati, non possono trovare alcun supporto nell'evidenza empirica. Oltre a tutto questo, le teorie che costituiscono le medicine alternative mancano di una dote particolarmente importante per giudicarne lo *status*: esse non sono in grado di fare previsioni precise. E poiché la capacità di prevedere i fenomeni caratterizza le teorie scientifiche autentiche, questa incapacità che caratterizza le medicine alternative costituisce un segno della loro debolezza scientifica.

Un'altra caratteristica delle medicine alternative è costituita dal fatto che le loro teorie non sono riformabili. Esse, infatti, sono sempre uguali a se stesse come gli antichi sistemi, e se da una teoria ne viene generata un'altra, questa non è considerata un'evoluzione della prima, ma una teoria rivale e alternativa.

Ma il vizio metodologico fondamentale di tutte le medicine eretiche risiede nel fatto che le loro teorie sono sempre inconfutabili. Se si analizzano i manuali delle varie dottrine prese in esame, si constata facilmente come *tutti* i fatti osservati vadano sempre d'accordo con le teorie proposte e come nessuna osservazione sia veramente capace di mettere in dubbio le assunzioni di fondo. Se qualche osservazione pare mettere in pericolo la teoria o qualcuno dei concetti che la sostengono, viene immediatamente avanzata una ipotesi *ad hoc* che immunizza la teoria contro ogni confutazione.

Questa situazione costituisce una netta violazione di quel principio di falsificabilità che, dopo Popper, rappresenta il criterio fondamentale di demarcazione della scienza dalla pseudoscienza.

A questi difetti metodologici si deve aggiungere la particolare situazione sociale delle medicine alternative. È facile infatti constatare che fra il loro mondo e quello della medicina scientifica manca ogni forma di comunicazione. Per quanto gli aderenti alle medicine eretiche accusino spesso il mondo della scienza di volerli tenere al bando, l'ostracismo di cui sono oggetto è dovuto esclusivamente al fatto che essi non si assoggettano alle regole rigorose che vigono nella comunità scientifica internazionale e che vengono accettate da tutti i ricercatori.

Le più importanti e diffuse fra le medicine alternative possiedono, per esempio, riviste proprie e gli aderenti pubblicano i loro lavori esclusivamente su queste riviste. Ovviamente ciò impedisce di fatto ogni reale esame e ogni controllo dei risultati ottenuti dai medici alternativi da parte della comunità scientifica. In tal modo ogni singola medicina alternativa si chiude nel suo mondo e, rifuggendo da ogni confronto con le idee che stanno al di fuori di essa, continua ad affermare le proprie verità, perpetuandole.

Come ha sottolineato Ziman, «tutt'intorno ai confini definiti con una relativa precisione della scienza istituzionalizzata "ufficiale" si sviluppa naturalmente una frangia di idee "parascientifiche", che cercano un riconoscimento formale dalla comunità scientifica o dalla gente in genere. Ma il parascientismo è per lo scienziato esperto un disordine pericoloso, che tende ad abbassare le sue difese scettiche e può suscitare una straordinaria disposizione alla credulità e all'illusione. Per coloro che ne sono colpiti, l'unica risposta della comunità scientifica deve essere: "Porta prove attendibili, porta valide argomentazioni e noi saremo disposti a farci convincere, ma, fino a quel giorno, non devi aspettarti da noi una grande fiducia nelle tue affermazioni, né un gran sostegno a una causa della quale non siamo veramente convinti"».

> **«Ogni volta che uno scienziato pretende che la sua teoria sia sostenuta dall'esperienza e dall'osservazione dovremmo porgli la domanda seguente: puoi descrivere una qualsiasi osservazione possibile che, effettivamente compiuta, confuterebbe la tua teoria? Se non lo puoi, allora è chiaro che la teoria non ha il carattere di una teoria empirica; infatti se tutte le osservazioni concepibili vanno d'accordo con la tua teoria, allora non hai il diritto di pretendere che una qualsiasi osservazione particolare offra un sostegno empirico alla tua teoria. Oppure, per dirla in breve: solo se puoi dirmi in qual modo la tua teoria possa essere confutata o falsificata, possiamo accettare la pretesa che la tua teoria abbia il carattere di una teoria empirica.»**
>
> Karl Popper
> da *Scienza e filosofia* (Einaudi, Torino, 1982)

Allegoria raffigurante il malato assistito da tre «specialisti»: medico, speziale e barbiere. La stampa è conservata presso l'Istituto di storia della medicina di Bologna.

BIBLIOGRAFIA

CARREL ALEXIS, *Medicina ufficiale e medicine eretiche*, Bompiani, Milano, 1950.

FEDERSPIL GIOVANNI, *I fondamenti del metodo in medicina clinica e sperimentale*, Piccin, Padova, 1980.

FEDERSPIL GIOVANNI e SCANDELLARI CESARE, *Medicina scientifica e medicina alternativa*, Parte I in «Medicina» Riv. E.M.I., 4, pp. 433-442, 1984; Parte II in «Medicina» Riv. E.M.I., 5, pp. 89-104, 1985, USES, Firenze.

CASSOLI PIERO e IANNUZZO GIOVANNI, *Ricerca sulla pranoterapia e sui guaritori. La pratica e i risultati valutati dalla scienza*, RED/Studio Redazionale, Como, 1988.

VELIMIROVIC BORIS, *Definizioni e funzioni della medicina alternativa* in «Nuova Civiltà delle Macchine», 4, pp. 56-69, 1988.

HAHNEMANN SAMUEL, *Organon. Dell'arte del guarire*, RED/Studio Redazionale, Como, 1989.

Gel «intelligenti»

Lo straordinario comportamento di certi aggregati di molecole a catena lunga, che sono in grado di gonfiarsi o di ridursi in risposta a vari stimoli, potrebbe costituire la base per macchine di tipo del tutto nuovo

di Yoshihito Osada e Simon B. Ross-Murphy

I prodotti industriali sono costituiti generalmente da sostanze per loro natura resistenti, dure, asciutte e facili da lavorare, come metalli, ceramiche o materie plastiche. In genere si evitano componenti «umidi», come liquidi e gel, in quanto i primi non hanno una forma propria e i secondi sono deboli e tendono a cedere anche sotto carichi minimi. In effetti, un gel può essere chimicamente instabile e le sue proprietà si deteriorano col disseccamento. Attualmente queste sostanze trovano impiego solo per applicazioni specializzate, quali ingredienti alimentari, materiali assorbenti e lenti a contatto morbide.

Tuttavia un numero crescente di studiosi, ispirandosi alla natura, ha iniziato a scorgere nuove opportunità in questi materiali. I sistemi biologici sono costituiti in massima parte da sostanze morbide e umide; in effetti esistono molti esseri viventi che mancano totalmente di struttura rigida. L'oloturia, per esempio, è essenzialmente un gel imbevuto d'acqua che contiene organi rudimentali; e tuttavia può nutrirsi, riprodursi e difendersi dai predatori. Essa reagisce rapidamente al tocco irrigidendo il corpo, solitamente flessibile, e se viene maneggiata con rudezza può emettere un fluido viscido che impedisce al disturbatore di afferrarla saldamente.

Per creare i cosiddetti sistemi biomimetici si impiegano materiali di tipo particolare, i gel polimerici. Nei laboratori universitari e industriali sono stati prodotti gel che possono mutare dimensioni e forma, convertendo così direttamente energia chimica in lavoro meccanico. Questi materiali potrebbero trovare impiego in tutti quei luoghi in cui l'energia per alimentare dispositivi più convenzionali è limitata o difficile da ottenere: nelle profondità dell'oceano, nello spazio o all'interno del corpo umano.

Le macchine basate sui gel costruite finora non sono ovviamente che un pallido riflesso dei sistemi di percezione e autoregolazione osservabili nell'oloturia, e a maggior ragione dell'intricata coreografia di muscoli, tendini e organi che si ammira negli organismi superiori. Ma riteniamo che in un futuro non troppo lontano si troverà un modo per costruire macchine «morbide» in grado di reagire all'ambiente in maniera «intelligente»; le ricerche in questo campo hanno già condotto ad applicazioni in chimica, ingegneria meccanica e medicina.

Come fece notare circa 70 anni fa Dorothy Jordan Lloyd, i gel sono «più facili da riconoscere che da definire». Si può affermare che essi consistono di due componenti: un liquido e una rete di lunghe molecole polimeriche che trattengono il liquido stesso conferendo così al gel un minimo di consistenza.

Per comprendere come si forma un gel, si consideri una soluzione contenente molecole di un tipico polimero sintetico, come il polistirene. Ciascuna molecola è costituita da circa 10 000 unità collegate di stirene, con un peso totale intorno al milione di dalton; completamente allungata, sarebbe lunga circa tre micrometri. Lo spazio occupato in realtà dalla molecola è però molto minore di quanto farebbe pensare questa lunghezza perché la catena è flessibile. In termini statistici, vi sono molti più stati in cui la catena è avvolta o aggrovigliata che non stati in cui è completamente distesa. Pertanto la distanza media fra un'estremità e l'altra è circa un decimo della lunghezza della molecola distesa.

In una soluzione diluita, ciascuna «matassa» polimerica è per lo più indipendente da tutte le altre, ma all'aumentare della concentrazione di polimero gli spazi occupati dalle varie molecole cominciano a sovrapporsi. Dato che l'attrito fra le molecole polimeriche impedisce un rapido flusso, la soluzione comincia a diventare densa e viscosa; a concentrazioni ancora più elevate le catene si intrecciano e si aggrovigliano l'una con l'altra come spaghetti. Questo tipo di sistema è viscoelastico: ha insieme le proprietà di una soluzione viscosa e di un solido elastico. Per esempio, se si potesse afferrare uno dei filamenti e tirarlo lentamente, esso fluirebbe fuori dalla massa attorcigliata delle altre catene; se invece si desse un rapido strattone a un singolo filamento, l'intera massa reagirebbe come un tutto unico.

Le definizioni di «lento» e «rapido» dipendono dal cosiddetto tempo di rilassamento delle catene molecolari in soluzione; questo valore è una misura del tempo impiegato dalla molecola a tornare in uno stato ripiegato dopo essere stata distesa, ed è determinato principalmente dalla massa della molecola in questione. Ogni molecola ha un suo intervallo di tempi di rilassamento e in una soluzione reale sono contenute molecole di differenti lunghezze; queste soluzioni mostrano perciò un'ampia distribuzione di tempi di rilassamento. Esse tendono gradualmente a comportarsi sempre più come solidi all'aumentare della velocità con cui sono sospinte e tirate. Anche la lunghezza e la rigidità delle catene polimeriche influiscono sul comportamento viscoelastico: entrambi i fattori accrescono il volume effettivo occupato da una singola catena e, di conseguenza, la probabilità che le catene interferiscano.

In certe circostanze le catene polimeriche in soluzione non si limitano a intrecciarsi; fra l'una e l'altra si formano legami permanenti che danno origine a grandi catene ramificate. Alla fine almeno una di queste megamolecole in soluzione arriva a estendersi da un lato all'altro del contenitore: è questa la transizione da soluzione viscoelastica a gel. Il campione gelificato non fluisce più come una soluzione polimerica, ma presenta le proprietà di un solido. Per esempio una perturbazione meccanica su un lato del campione si propaga lungo le catene molecolari fino al lato opposto.

In alcuni gel i legami che tengono unita la rete molecolare sono legami covalenti classici, in cui due atomi condividono una coppia di elettroni, ma vi sono anche configurazioni di legame meno rigide. I legami crociati possono essere prodotti da forze di van der Waals, che si esercitano tra atomi adiacenti, da interazioni idrofobe, che avvicinano parti delle molecole che tendono a sfuggire l'acqua, e da legami a idrogeno, in cui due molecole sono tenute unite per mezzo di un atomo di idrogeno.

Qualche volta questi legami non sono localizzati, come accade per i legami covalenti, ma formano invece «zone di giunzione» estese. A un livello ancora più complesso, i polimeri biologici sono tenuti insieme da interazioni tra strutture molecolari a grande scala. Per esempio, i dolcetti di gelatina - che sono a base di gelatina animale derivata dalle ossa - devono la loro tipica consistenza all'intreccio di una triplice elica proteica.

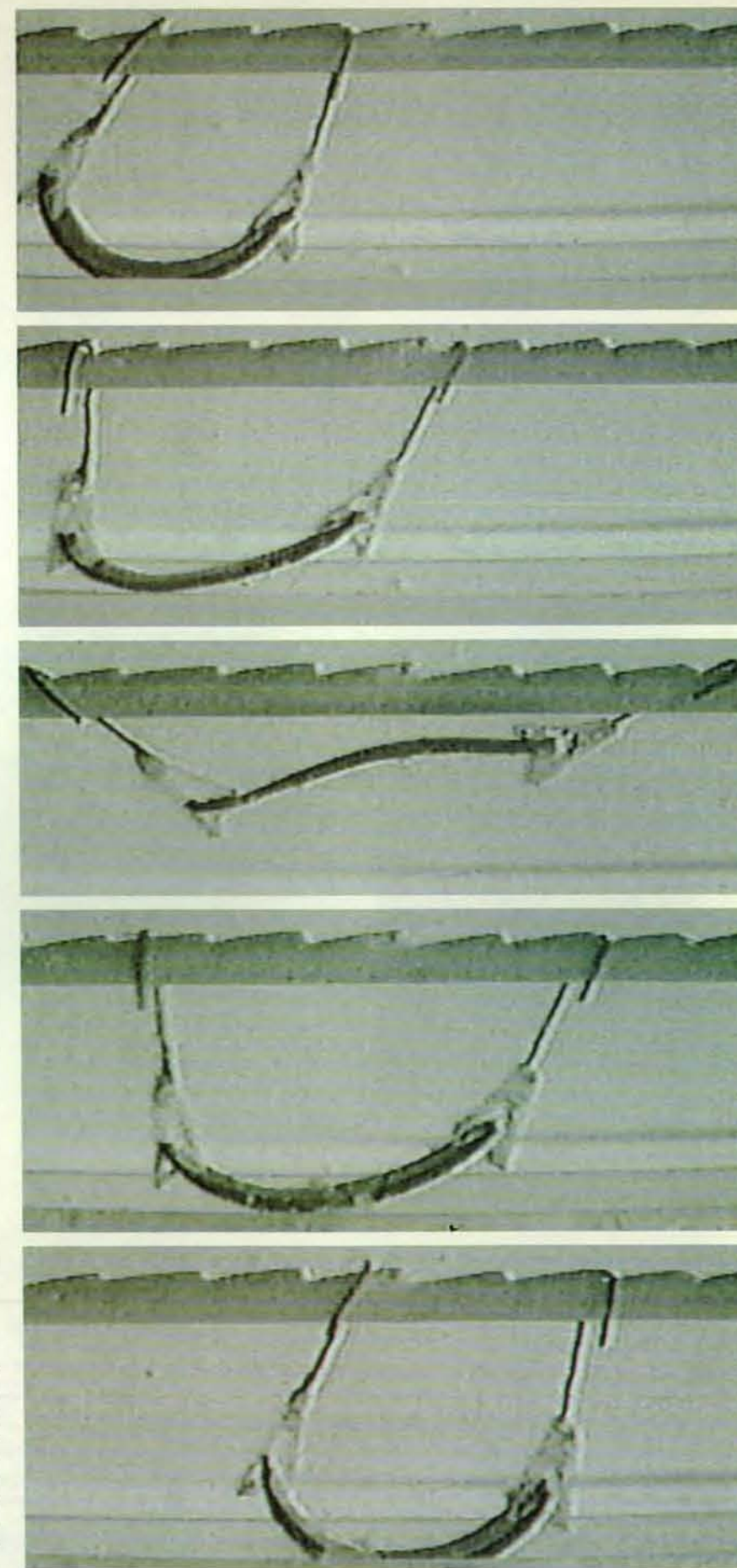
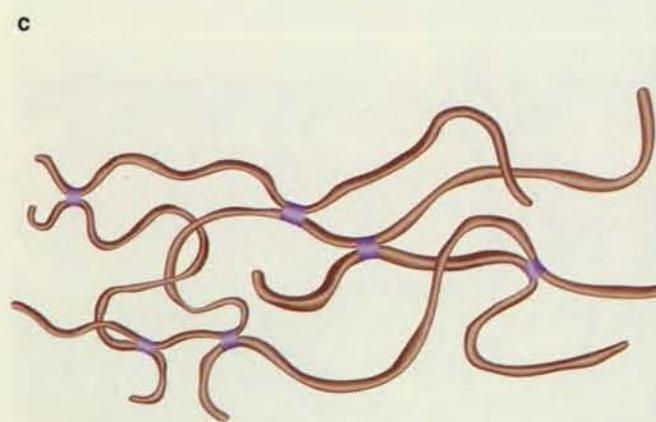
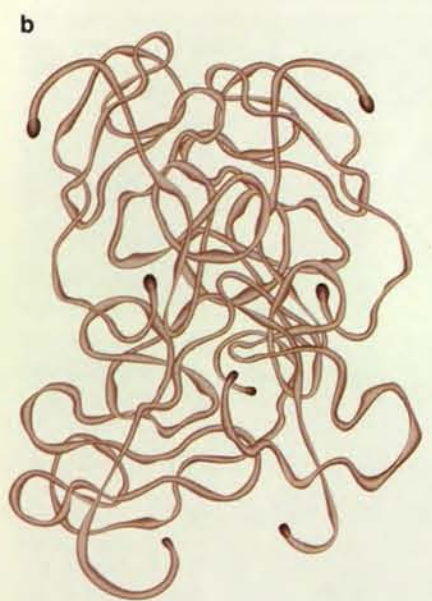
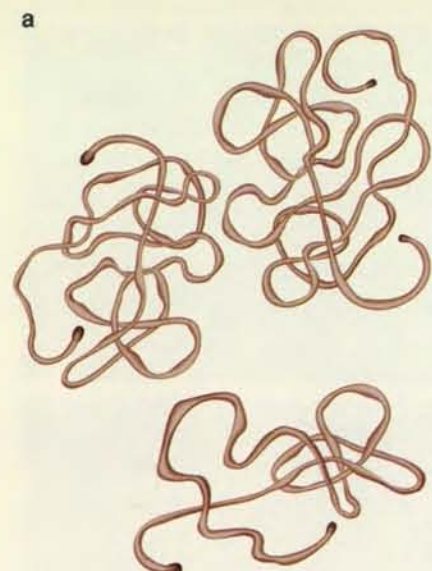
La rete aggrovigliata di filamenti polimerici non è che un aspetto della natura essenziale di un gel; un secondo aspetto, egualmente importante in linea di principio e ampiamente predominante in volume, è il solvente che pervade la rete e le dà corpo. La quantità di solvente contenuta in un dato gel dipende da una complessa interazione fra l'elasticità della rete polimerica e l'affinità degli atomi di questa per quelli del solvente.

Una singola molecola di polimero posta in un solvente può distendersi o raccogliersi su se stessa. Se le componenti del polimero e gli atomi del solvente si attraggono a vicenda, le dimensioni medie delle spire della molecola tendono ad aumentare per massimizzare il numero di interazioni fra polimero e molecole del solvente. Il polimero si gonfia e il solvente è definito «termodinamicamente buono». Viceversa, in un cattivo solvente le componenti del polimero tendono a sfuggire le molecole del solvente e il polimero sembra contrarsi.

Se un gel viene posto in un buon solvente, le sue dimensioni aumentano finché la rete polimerica si stira al punto che la forza elastica risultante si contrappone all'afflusso di molecole di solvente. La forza elastica dipende dalla quantità di legami crociati (che limitano la massima estensione delle catene polimeriche); pertanto un gel con numerosi legami di questo tipo si rigonfia meno di uno che ha pochi legami crociati.

Se il polimero di cui è costituito un gel contiene nello scheletro gruppi dotati di carica (molecole che accettano o do-

Questa struttura vermiforme di gel, che si muove incurvandosi e raddrizzandosi ripetutamente, è stata realizzata da Osada. Molecole di tensioattivo nel liquido che circonda la struttura si raccolgono sulla superficie superiore del gel per effetto di un campo elettrico, facendo sì che il gel si contragga. Quando si varia la polarità del campo elettrico, il tensioattivo ritorna in soluzione.

a

Le molecole a catena lunga (*a*) fanno sì che una soluzione diventi viscosa perché interferiscono l'una con l'altra mentre la soluzione scorre. All'aumentare della concentrazione, le molecole si aggrovigliano, dando origine a un comportamento viscoelastico che unisce caratteristiche tipiche dei solidi e dei liquidi (*b*). Se le molecole intrecciate formano legami, ciò che ne risulta è un gel (*c, d*).

b

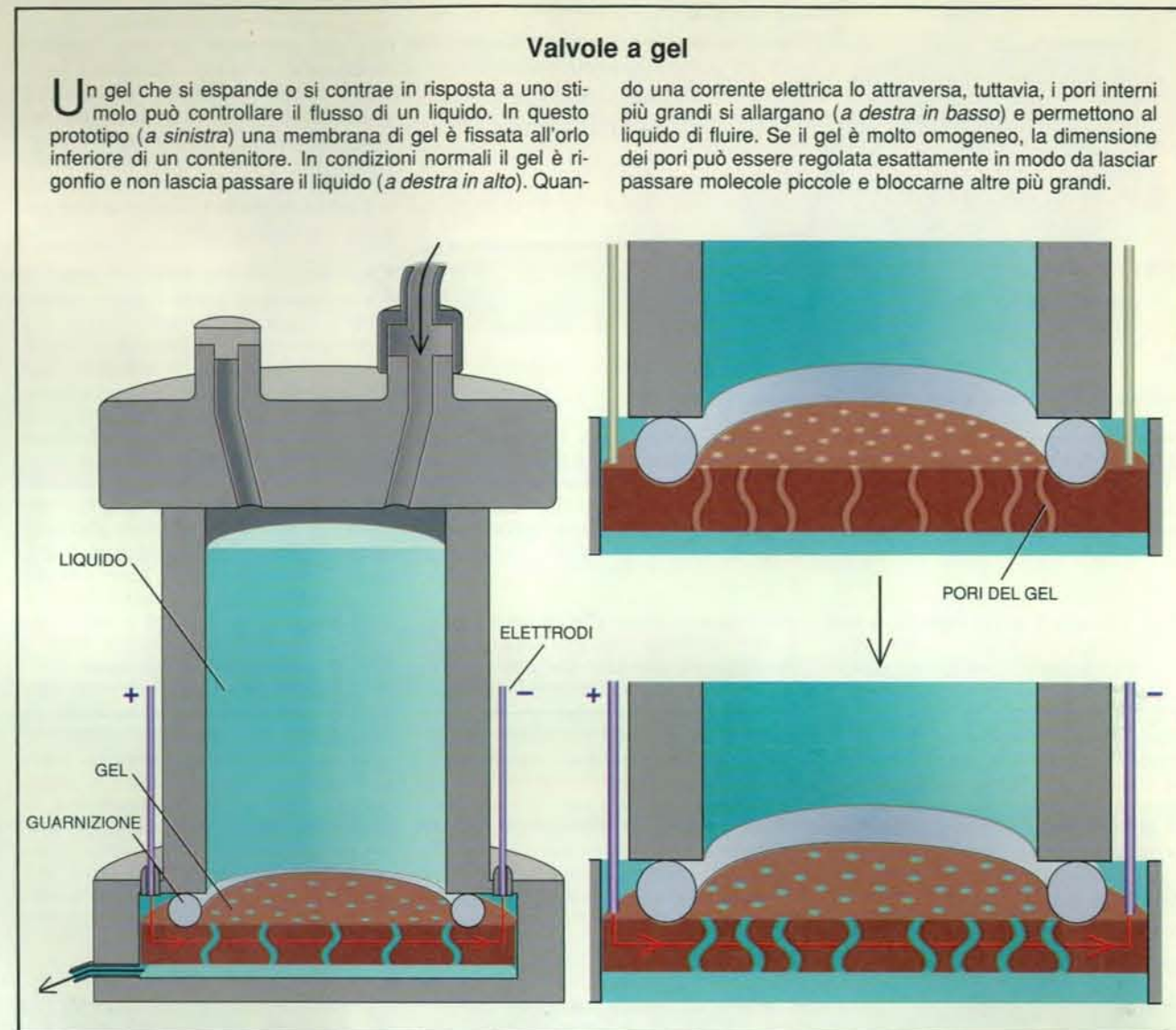
nano facilmente elettroni), allora possono entrare in gioco ulteriori effetti. Il primo di questi è il cosiddetto effetto del polielettrolita. In acqua pura, un polimero che contiene gruppi dotati di carica tenderà ad aumentare le proprie dimensioni allo scopo di minimizzare la repulsione dei gruppi stessi. Se si discioglie in acqua un elettrolita semplice, come il sale da cucina, gli ioni di carica opposta a quelli presenti sul polimero possono neutralizzare la carica di quest'ultimo. Così, all'aumentare della forza ionica, il polimero riprende la propria forma ripiegata. Un gel di polielettrolita si rigonfia enormemente in acqua pura o in soluzioni elettrolitiche di bassa forza ionica, mentre si contrae in presenza di alte concentrazioni di elettroliti. L'effetto può essere intensificato regolando il «cocktail» di specie ioniche in modo da massimizzare la schermatura.

La maggior parte dei gel si gonfia o si contrae in misura strettamente proporzionale alla qualità termodinamica del solvente che li pervade. Alcuni, tuttavia, subiscono una variazione improvvisa di dimensioni in risposta a una modificazione relativamente piccola della qualità del solvente. Per esempio, se un gel di poliacrilammide parzialmente carico è immerso in una miscela di etanolo e acqua, si contrae lievemente via via che si aggiunge etanolo alla soluzione, fino al punto in cui l'aggiunta di una minuscola quantità di etanolo fa sì che il gel si riduca a una piccola frazione del suo volume precedente. Questo fenomeno è analogo al comportamento di un «fluido critico» (come l'anidride carbonica ad alta pressione) vicino alla transizione di fase liquido-vapore. In questa situazione una variazione molto piccola di temperatura o pressione può convertire il liquido nella fase vapore o viceversa.

L'improvvisa contrazione del gel di poliacrilammide dipende da un'interazione che si evolve a poco a poco fra l'affinità della rete per le molecole del solvente e le forze elastiche che la tengono unita. Dapprima il gel si contrae lentamente, ma, via via che espelle solvente, le sue catene molecolari interagiscono a vicenda con più forza; questa interazione tende a espellere altro solvente e quindi ad autorafforzarsi, fino al punto in cui la rete si è ridotta in uno stato strettamente ripiegato. Se si aggiunge altra acqua, riducendo così la concentrazione di etanolo, il ciclo di retroazione descritto si svolge alla rovescia, e il gel riprende le proprie dimensioni iniziali.

In dipendenza dalla struttura esatta del gel, questo semplice svolgimento può essere complicato da un certo numero di fattori; in alcuni casi il gel funge da membrana semipermeabile e l'interazione fra ioni del solvente e siti di carica sul gel diventa il fattore più importante nella transizione fra contrazione e rigonfiamento. Oltre a ciò, la nettezza della transizione dipende anche dalla rigidità delle catene polimeriche del gel. Sia le molecole flessibili sia quelle estremamente rigide tendono a dar luogo a transizioni continue.

Negli ultimi dieci anni studiosi di tutto il mondo hanno messo a punto nuovi gel che si gonfiano o si contraggono in risposta a molti stimoli differenti - temperatura, *p*H o campi elettrici - a seconda della composizione chimica del gel e del solvente.

Un gel derivato dal poli-N-isopropilacrilammide può ridursi al 30 per cento del suo volume iniziale quando viene riscaldato al di sopra di una temperatura critica. Un gel analogo, ottenuto inducendo la formazione di legami crociati in una soluzione acquosa di polimetilviniletere, va incontro a rigonfiamento e contrazione rapidi e reversibili a 37 gradi Celsius. Il diametro delle fibrille di gel passa da 400 micrometri a 20 gradi Celsius a 200 micrometri a 40 gradi.

c

Valvole a gel

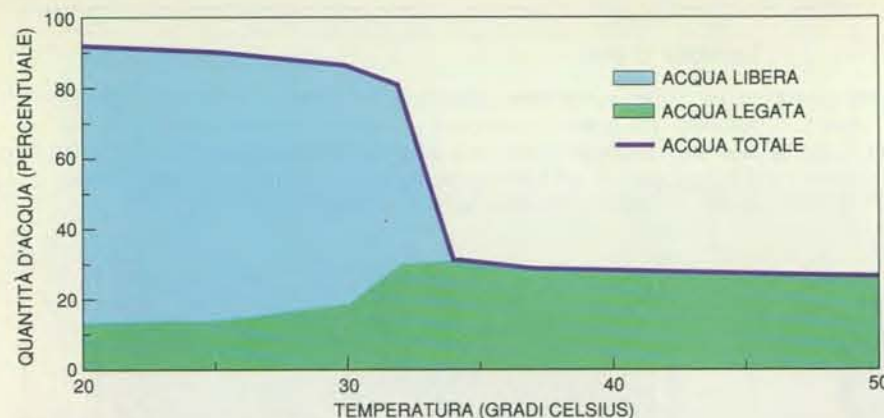
Un gel che si espande o si contrae in risposta a uno stimolo può controllare il flusso di un liquido. In questo prototipo (*a sinistra*) una membrana di gel è fissata all'orlo inferiore di un contenitore. In condizioni normali il gel è rigonfio e non lascia passare il liquido (*a destra in alto*). Quando una corrente elettrica lo attraversa, tuttavia, i pori interni più grandi si allargano (*a destra in basso*) e permettono al liquido di fluire. Se il gel è molto omogeneo, la dimensione dei pori può essere regolata esattamente in modo da lasciar passare molecole piccole e bloccarne altre più grandi.

Un campo elettrico di non più di mezzo volt per centimetro induce una analoga contrazione in un gel di poliacrilammide immerso in acetone e acqua. Le particelle del gel si gonfiano o si riducono a una velocità che dipende dalla corrente che fluisce attraverso di esse e dal quadrato del loro diametro. In teoria, un campo elettrico di cinque volt per millimetro potrebbe far contrarre in un millisecondo un insieme di particelle di gel di un micrometro di diametro, portandole fino al 4 per cento del volume originario. Questa reazione rapida potrebbe rendere i gel adatti a essere impiegati come «muscoli» in robot o altri dispositivi meccanici, o anche in protesi da applicare all'uomo.

Nella loro prima applicazione ad ampia scala, i gel «intelligenti» saranno probabilmente usati non come sostituti dei muscoli, ma come discendenti altamente perfezionati dei tradizionali cataplasmi. Trovare un modo per far arrivare in ogni dato momento una sostanza farmacologicamente attiva agli organi che ne hanno bisogno nelle dosi richieste è da molto tempo un problema cruciale in medicina. In anni recenti l'industria farmaceutica ha adottato membrane semipermeabili per liberare farmaci a velocità costante; sistemi basati sui gel «intelligenti» costituirebbero un progresso in questa direzione perché i gel potrebbero percepire le condizioni esistenti all'interno dell'organismo e variare la velocità di liberazione del farmaco in modo da mantenerne un livello appropriato nel circolo sanguigno.

Ronald A. Siegel e colleghi dell'Università della California a San Francisco hanno messo a punto un semplice sistema basato sui gel per proteggere sostanze sensibili agli acidi nell'ambiente gastrico. Il gel si contrae quando il *p*H è basso, ma nell'ambiente più alcalino dell'intestino si espande e diventa permeabile, consentendo al farmaco incapsulato di diffondere.

Sono stati anche ideati gel che liberano farmaci o biomolecole in risposta a variazioni di un campo elettrico. L'insulina può diffondere da un gel di polielettrolita contenente pochi legami crociati quando fluisce corrente elettrica, ma la diffusione cessa non appena la corrente è interrotta. Il gel potrebbe così costituire la base di una pompa impiantabile per insulina senza parti in movimento. Un gruppo di ricerca dell'Università di Trondheim, in Norvegia, e del Veterans Administration Islet Transplant Center di Los Angeles ha addirittura evitato di ricorrere al meccanismo a pompa racchiudendo cellule insulino-secernenti in un gel di alginato. Si spera di iniziare già nel corso di quest'anno la sperimentazione clinica sull'uomo.

Gel che si espandono e si contraggono per effetto della corrente elettrica possono anche fungere da «valvole chimiche» di impiego generale. Il gel è modellato in forma di membrana porosa fissata ai bordi. Quando esso si contrae, i pori della membrana sono costretti a espandersi, permettendo a liquidi e molecole in soluzione di attraversare la membrana

FONTE: Liang C. Dong e Allan S. Hoffman, Università di Washington

Una transizione di fase provoca, all'aumentare della temperatura, il collasso di un gel polimerico termosensibile. Il grafico rappresenta il contenuto d'acqua del gel in funzione della temperatura. Quando il gel è freddo, lo stato rigonfio è termodinamicamente stabile, ma, via via che la sua temperatura aumenta, interazioni molecolari stabilizzano invece lo stato in cui le catene sono ripiegate e aggrovigliate.

stessa. Quando il gel si espande, i pori si restringono e il flusso termina. Mantenendo la corrente a livelli intermedi si può controllare esattamente la dimensione dei pori e quindi determinare quali molecole possono attraversarli. Oggi le membrane regolabili vengono impiegate per separare miscele di solventi le cui molecole hanno dimensioni diverse.

I gel possono essere utilizzati anche per estrarre grandi molecole da soluzioni acquose diluite, un processo di possibile utilità industriale. Edward L. Cussler, Jr., e colleghi dell'Università del Minnesota hanno scoperto che un gel rigonfio assorbe preferenzialmente acqua, escludendo però le sostanze disciolte in essa. Il procedimento consiste nell'immergere semplicemente il gel e lasciare che si inzuppi, separando così il prodotto desiderato. Il gel viene poi fatto contrarre per espellere buona parte del liquido e quindi riportato allo stato iniziale per proseguire il processo di separazione.

Benché il rigonfiamento e la contrazione di gel fissati a un substrato si siano finora dimostrati le proprietà più utili, da oltre quarant'anni i ricercatori in questo campo sono attratti dall'idea di utilizzare i gel per produrre forza motrice. Aharon Katchalsky del Weizmann Institute of Science di Rehovot, in Israele, e Werner Kuhn dell'Università di Basilea hanno costruito i primi sistemi «chemiomeccanici» nel 1950, alterando il *p*H dell'ambiente acido in cui erano immersi i gel in modo da provocarne l'espansione e la contrazione.

Recentemente uno di noi (Osada) e collaboratori hanno messo a punto un nuovo sistema chemiomeccanico che è stato presentato nel dicembre 1991 al Second Polymer Gel Symposium and Robo-bug Fest di Tsukuba; altre otto organizzazioni giapponesi per la ricerca hanno esposto in quell'occasione dispositivi mobili chemiomeccanici.

Il nostro sistema consiste di una striscia vermiforme di gel che si sposta, per effetto di un campo elettrico alternato, lungo una bacchetta di sostegno. Il gel, appeso alla bacchetta per mezzo di uncini metallici, è immerso in una soluzione acquosa contenente un tensioattivo. Elettrodi disposti parallelamente sopra e sotto il gel ne controllano il moto.

Quando si applica agli elettrodi una differenza di potenziale, le molecole di tensioattivo, che hanno carica positiva, migrano verso l'elettrodo negativo. Lungo il cammino esse incontrano la superficie del gel carica negativamente e vi aderiscono, facendo sì che il gel si contragga. Le molecole di tensioattivo si posano preferenzialmente sulla faccia del gel che guarda verso l'elettrodo positivo; di conseguenza la striscia si contrae soprattutto su quel lato e si incurva verso l'alto. Quando si inverte la polarità del campo elettrico, le molecole di tensioattivo si staccano e il gel si distende nuovamente. Una serie di tacche a dente di sega sulla parte superiore della bacchetta di sostegno garantisce che la striscia di gel si muova in avanti a ogni ciclo di piegamento e distensione.

Per quanto semplice possa essere, questo sistema mostra tutte le caratteristiche essenziali che differenziano i sistemi chemiomeccanici «morbidi» dai dispositivi meccanici realizzati con materiali più rigidi. Contrariamente ai motori e alle pompe convenzionali, i gel sono soffici e flessibili e il loro movimento ricorda più quello dei muscoli che non quello di macchine fatte di metallo. Di solito questi moti fluidi si osservano solo nei sistemi biologici, come le ali degli uccelli che si riconfigurano continuamente così da massimizzare la portanza.

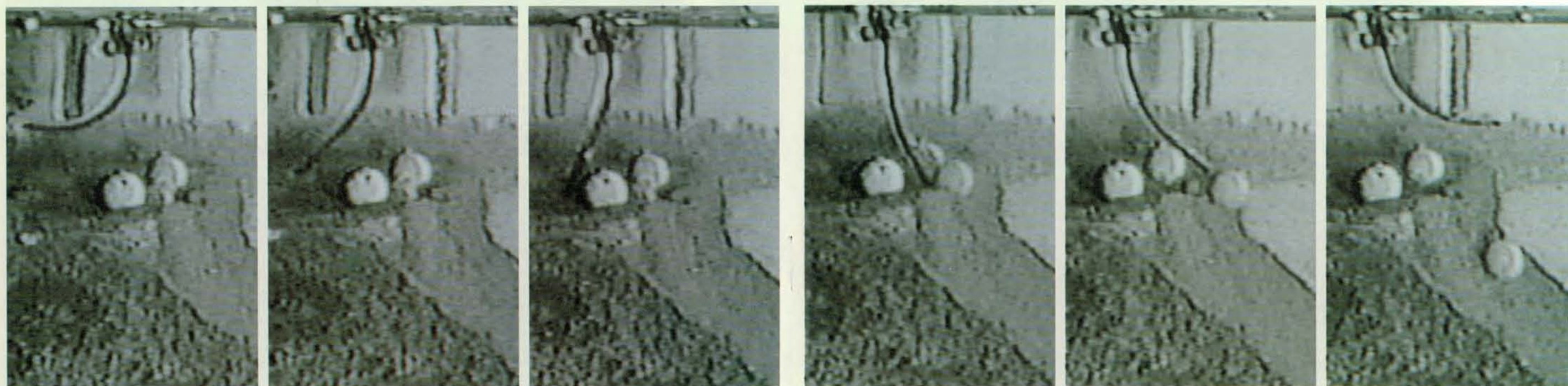
Essendo morbidi, i gel possono toccare materiali delicati senza danneggiarli. Ancora più importante, tuttavia, è il fatto che i gel sono flessibili rispetto al loro ambiente. Le macchine costituite da metallo o silicio funzionano come sistemi chiusi: non si adattano a variazioni delle condizioni operative a meno che non vengano controllate da un sistema separato di sensori o da un operatore umano. I gel, viceversa, sono termodinamicamente «aperti», in quanto scambiano sostanze chimiche con il solvente che li circonda e alterano il proprio stato molecolare nel compiere lavoro. Se si aggiunge al sistema energia libera sotto forma di nuove sostanze, i sistemi chemiomeccanici tornano al proprio stato originario senza ulteriori interventi. Noi riteniamo che un giorno sarà possibile sfruttare queste proprietà per creare macchine capaci di percezione e regolazione autonoma, che rispondano in modo intelligente alle variazioni dell'ambiente. Anche se le macchine «morbide» non sostituiranno mai quelle tradizionali, fra breve il termine *wetware* potrebbe trovar posto accanto ad hardware e software nel lessico ingegneristico.

La contrazione asimmetrica è il fenomeno che determina il comportamento dei gel fotografati in questo articolo. L'incurvamento del gel è provocato da molecole di tensioattivo che migrano nel solvente per effetto di un campo elettrico. Quando si applica il campo, le «teste» delle molecole, cariche positivamente, si dirigono verso i siti di carica negativa del gel; così facendo allontanano le molecole d'acqua e provocano la contrazione di una faccia del gel. Quando il campo viene invertito le molecole di tensioattivo diffondono di nuovo nella soluzione e il gel si distende.

BIBLIOGRAFIA

TANAKA TOYOICHI, *I gel* in «Le Scienze» n. 151, marzo 1981.

BURCHARD W. e ROSS-MURPHY S. B. (a cura), *Physical Networks: Polymers and Gels*, Elsevier Science Publishers, 1990.

DEROSSI D., KAJIWARA K., OSADA Y. e YAMAUCHI A., *Polymer Gels: Fundamentals and Biomedical Applications*, Plenum Press, 1991.

Questa serie di immagini mostra la capacità di un gel «intelligente» di interagire con l'esterno. Una striscia di gel, fatta con lo stesso materiale di quella a pagina 45, si piega da un lato e poi dall'altro per effetto di un campo elettrico. Così facendo colpisce una sferetta e la spinge giù per un pendio. Il materiale della «mazza», pur essendo abbastanza robusto da spostare la sfera, deve essere immerso in un liquido, e quindi nell'impiego pratico sarebbe probabilmente racchiuso in un contenitore.

Come le cellule reagiscono allo stress

In condizioni di emergenza le cellule, per riparare i danni, sintetizzano proteine «da stress»: si spera che la conoscenza del loro meccanismo d'azione permetta di combattere infezioni, malattie autoimmuni e cancro

di William J. Welch

Immediatamente dopo un improvviso aumento della temperatura, tutte le cellule - dal più semplice batterio al neurone più specializzato - incrementano la produzione di una certa classe di molecole che servono a tamponare i danni subiti. Quando, trent'anni or sono, questo fenomeno venne osservato per la prima volta, lo si definì «risposta allo shock termico». Studi successivi hanno rivelato che la stessa reazione si verifica quando le cellule subiscono una grande varietà di attacchi ambientali, tra cui intossicazioni da metalli, alcoli e molte altre tossine. Il fenomeno ha luogo in cellule coltivate *in vitro* che vengono sottoposte a un trauma, nei tessuti di bambini febbricitanti e negli organi di pazienti colpiti da infarto cardiaco o sottoposti alla chemioterapia per il cancro. Dato che stimoli così numerosi e di natura diversa attivano uno stesso meccanismo di difesa cellulare, questo viene oggi comunemente indicato come «risposta allo stress» e le proteine che in esso vengono espresse sono definite «proteine da stress».

Nell'analizzare la struttura e la funzione delle proteine da stress, ci si rese conto che si trattava di qualcosa di ben più importante di semplici molecole difensive. Durante tutta la vita di una cellula, molte di queste proteine partecipano a processi metabolici essenziali, compresi quelli che portano alla sintesi e all'assemblaggio di tutte le altre proteine cellulari. Sembra che alcune proteine da stress orchestrino l'attività di molecole regolatrici della crescita e del differenziamento cellulari.

La natura di queste proteine è ancora in parte oscura, ma si cominciano già a trovare alcuni modi per fare buon uso della risposta allo stress, che mostra notevoli potenzialità nel monitoraggio dell'inquinamento e nel miglioramento dei test tossicologici. Le applicazioni mediche nella lotta contro le infezioni, il cancro e le malattie immunitarie costituiscono forse un traguardo più lontano, ma chiaramente visibile all'orizzonte.

Tutte queste possibilità non potevano certo essere previste da quei ricercatori che per primi scoprirono la risposta allo stress. Come capita spesso nella scienza, fu una scoperta casuale, del tutto inattesa. Agli inizi degli anni sessanta, i biologi che studiavano la base genetica dello sviluppo animale erano soliti concentrare la propria attenzione sul moscerino della frutta (*Drosophila melanogaster*). Era questo un organismo adatto per studiare la maturazione da embrione a individuo adulto, in parte a causa di un'insolita caratteristica genetica. Le cellule delle ghiandole salivari hanno quattro cromosomi nei quali la dotazione normale di DNA è duplicata migliaia di volte; tutte le copie sono allineate l'una a fianco dell'altra. Questi cromosomi, cosiddetti politenici, sono così grandi da essere visibili al microscopio ottico. Durante ogni stadio del processo di sviluppo, regioni distinte di questi cromosomi si ingrossano; ogni rigonfiamento è il risultato di un cambiamento specifico nell'espressione di un gene.

Nel corso dei suoi studi presso l'Istituto internazionale di genetica e biofisica di Napoli Ferruccio M. Ritossa notò che, in seguito all'esposizione di ghiandole salivari isolate a temperature leggermente al di sopra di quelle ottimali per la crescita e lo sviluppo di *Drosophila*, la formazione dei rigonfiamenti nei cromosomi seguiva uno schema diverso. Il fenomeno si manifestava entro un minuto o due dall'aumento di temperatura, e i rigonfiamenti continuavano poi a ingrossarsi anche per 30-40 minuti. Nel decennio successivo, ai risultati di Ritossa si aggiunsero quelli di altri ricercatori.

I vermi prodotti per manipolazione genetica sono di solito incolori (*qui a lato*), ma possono diventare blu-verdi (*pagina a fronte*) quando sono esposti a tossine, a calore eccessivo o ad altre condizioni avverse. La colorazione è provocata dall'attività di un gene indicatore, associata all'espressione di geni per le proteine da stress, che aiutano gli organismi a far fronte a condizioni di vita sfavorevoli.

I rigonfiamenti nei cromosomi politenici di *Drosophila melanogaster* indicano la presenza di un'attività genica locale. A mano a mano che *Drosophila* attraversa i diversi stadi di sviluppo, la distribuzione dei rigonfiamenti si modifica. Anche temperature eccezionalmente elevate stimolano la formazione di certi rigonfiamenti, come si osserva nell'illustrazione qui a fianco. I rigonfiamenti indicano l'espressione dei geni per le proteine da shock termico che appartengono alla famiglia molecolare hsp 70.

I livelli delle proteine da shock termico nelle cellule aumentano a mano a mano che cresce la temperatura. Nei gel elettroforetici illustrati qui sopra, ogni banda orizzontale corrisponde a una proteina presente nelle cellule di *Drosophila*. Con l'aumentare della temperatura, le cellule cessano di produrre la maggior parte delle proteine e cominciano invece a sintetizzare una rilevante quantità di proteine da shock termico. Le più diffuse tra queste appartengono alla famiglia hsp 70, che comprende molecole con un peso attorno ai 70 000 dalton.

Nel 1974 Alfred Tissières, ricercatore ospite dell'Università di Ginevra, e Herschel K. Mitchell del California Institute of Technology dimostrarono che la formazione di rigonfiamenti indotta dal calore si associa all'espressione di livelli elevati di particolari proteine «da shock termico». Questi rigonfiamenti cromosomici rappresentano siti della molecola di DNA in cui vengono prodotte specifiche molecole di RNA messaggero. L'RNA messaggero contiene l'informazione genetica necessaria per sintetizzare le singole proteine da shock termico.

Alla fine degli anni settanta erano già numerosi i dati che dimostravano come la risposta a uno shock termico è una proprietà generale di tutte le cellule. In seguito a un improvviso aumento della temperatura, cellule batteriche, di lievito, vegetali e animali coltivate *in vitro* esprimevano tutte una quantità superiore al normale di proteine che erano simili, per dimensione, a quelle da shock termico prodotte dai moscerini della frutta. Inoltre si era anche osservato che le cellule sintetizzavano una o più proteine da shock termico ogniqualvolta venivano esposte a metalli pesanti, alcoli e varie altre tossine metaboliche.

Poiché stimoli tossici tanto numerosi e diversi portavano tutti a cambiamenti simili dell'espressione genica, la risposta allo shock termico ha cominciato a essere definita, più in generale, risposta allo stress, e i prodotti a essa associati sono stati chiamati proteine da stress. Si è anche cominciato a pensare che questa risposta estremamente generale a cambiamenti ambientali avversi rappresenti un fondamentale meccanismo cellulare di difesa. Le proteine da stress, che a quanto pare sono espresse soltanto in caso di difficoltà, fanno presumibilmente parte di questa risposta.

Negli anni immediatamente successivi, è divenuto sempre più chiaro che le proteine da stress svolgono effettivamente un ruolo attivo nella difesa cellulare. Si è riusciti a identificare e a isolare i geni che codificano per singole proteine da stress e si è visto che mutazioni in questi geni producono anomalie cellulari interessanti. Per esempio, i batteri con mutazioni a carico dei geni che codificano per parecchie proteine da stress presentano difetti nella sintesi di DNA e di RNA, non hanno la capacità di dividersi normalmente e appaiono incapaci di degradare le proteine in maniera corretta. Inoltre questi mutanti non riescono a crescere a temperature elevate.

Ben presto si è visto che, come avviene nei batteri, la risposta allo stress ha un ruolo importante anche nella capacità delle cellule animali di resistere a brevi esposizioni a temperature elevate. Le cellule animali sottoposte a un lieve shock termico - sufficiente a far aumentare i livelli delle proteine da stress - risultano meglio protette da un secondo trattamento termico, che sarebbe altrimenti letale. Inoltre le cellule che tollerano il calore sono anche meno suscettibili ad altri agenti tossici. Si è giunti alla conclusione che la risposta allo stress protegge in qualche modo le cellule da insulti ambientali di varia natura.

Mentre si continuava a isolare da diversi organismi e a caratterizzare i geni che codificano per le proteine da stress, sono emersi due risultati inattesi. Innanzitutto, molti geni che codificano per le proteine da stress sono notevolmente simili in tutti gli organismi. Elizabeth A. Craig e collaboratori dell'Università del Wisconsin hanno riferito che i geni per la proteina da shock termico 70 (hsp 70), che è la proteina da stress indotta con maggiore frequenza, sono per più del 50 per cento identici nei batteri, nei lieviti e nel moscerino della frutta. È chiaro, dunque, che le proteine da stress devono essersi conservate lungo tutto il corso dell'evoluzione e probabilmente hanno svolto una funzione analoga e importante in tutti gli organismi.

Il secondo risultato inatteso è stato che molte proteine da stress si esprimono anche in cellule normali e non soltanto in cellule sottoposte a stress o traumatizzate. Di conseguenza queste proteine sono state suddivise in due gruppi: quelle costitutive, espresse in condizioni di crescita normale, e quelle indotte soltanto in cellule sottoposte a stress.

Sussistevano ancora perplessità su come tanti stimoli tossici apparentemente diversi potessero condurre sempre a un incremento dell'espressione dello stesso gruppo di proteine. Nel 1980 Lawrence E. Hightower, che lavorava all'Università del Connecticut, fornì una risposta plausibile. Egli notò che molti agenti che inducevano risposte allo stress erano denaturanti proteici, ossia sostanze che fanno perdere alle proteine la loro configurazione. Una proteina consiste in lunghe catene di amminoacidi avvolte in modo da assumere una ben precisa conformazione. Un'alterazione della struttura può portare a una perdita della funzionalità biologica della proteina stessa.

Pertanto Hightower propose che l'accumulo nella cellula di proteine denaturate o avvolte in modo anormale scateni una risposta allo stress; a suo parere, le proteine da stress possono in qualche modo facilitare l'identificazione e la rimozione delle proteine denaturate dalla cellula traumatizzata. Nel volgere di pochi anni Richard Voellmy dell'Università di Miami e Alfred L. Goldberg della Harvard University verificarono e confermarono l'ipotesi di Hightower: in uno studio fondamentale, essi dimostrarono che l'iniezione di proteine denaturate in cellule viventi basta per indurre una risposta allo stress.

Da allora in poi, in parecchi laboratori si sono cominciate a purificare le proteine da stress e a caratterizzarne le proprietà biochimiche. La hsp 70, la proteina da shock termico più facilmente inducibile, è stata oggetto della maggior parte di queste ricerche. Con l'aiuto di sonde molecolari si è visto che questa proteina si accumula in gran parte all'interno di una struttura nucleare, il nucleolo; questo produce i ribosomi, gli organelli che sintetizzano le proteine. La localizzazione dell'hsp 70 era interessante: ricerche precedenti avevano dimostrato che, in seguito a shock termico, le cellule cessano di produrre ribosomi e al contrario, nel loro nucleolo, si accumula una gran quantità di particelle ribosomali denaturate. Pertanto Hugh R. B. Pelham del Laboratory of Molecular Biology del Medical Research Council a Cambridge, in Inghilterra, ha proposto che l'hsp 70 possa in qualche modo riconoscere le proteine intracellulari denaturate e reintegrarle nella conformazione correttamente avvolta, dotata di attività biologica.

Nel 1986 Pelham e il suo collaboratore Sean Munro sono riusciti a isolare parecchi geni che codificano tutti per proteine affini alla hsp 70 e hanno notato che una forma di questa proteina è identica alla proteina che si lega alle immunoglobuline (BiP). In precedenza altri ricercatori avevano dimostrato che la BiP è attiva nella preparazione delle immunoglobuline - o anticorpi - e di altre proteine per la secrezione. Essa si lega alle proteine appena sintetizzate mentre queste si avvolgono o si assemblano nella loro conformazione matura. Se le proteine non riescono a compiere correttamente questo processo, rimangono legate alla BiP e alla fine vengono degradate. Inoltre, in condizioni in cui si forma un accumulo eccessivo di proteine avvolte in modo anomalo, la cellula sintetizza una maggiore quantità di BiP.

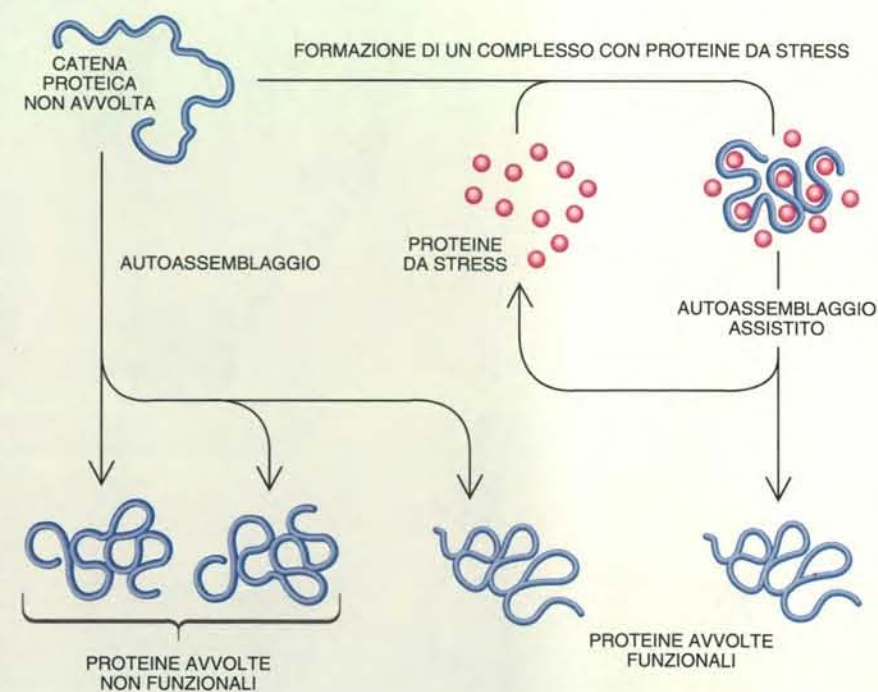
Queste osservazioni, considerate nel loro complesso, indicano che la BiP prende parte alla regolazione delle prime fasi della secrezione proteica. Sembra che essa agisca come un supervisore molecolare del controllo di qualità, consentendo alle proteine avvolte nella maniera corretta di accedere alla via della secrezione e trattenendo invece quelle proteine che sono incapaci di avvolgersi correttamente.

A mano a mano che veniva alla luce un numero sempre più elevato di geni che codificano per proteine analoghe all'hsp 70 e alla BiP, ci si è resi conto che esiste un'intera famiglia di proteine affini all'hsp 70. Tutte queste proteine hanno in comune certe proprietà, tra cui una grandissima affinità per l'adenosintrifosfato (ATP), la molecola che funge universalmente da fonte di energia intracellulare. Con una sola eccezione, tutte queste proteine affini sono presenti nelle cellule che si sviluppano in condizioni normali (sono cioè proteine costitutive), ma nelle cellule che subiscono stress metabolico esse vengono sintetizzate a livelli molto più elevati. Inoltre tutte mediano la maturazione di altre proteine cellulari, proprio come fa la BiP. Per esempio, le forme citoplasmatiche dell'hsp 70 interagiscono con molte altre proteine sintetizzate dai ribosomi.

In cellule sane, o comunque non soggette a stress, l'interazione della proteina affine all'hsp 70 con le proteine immature è transitoria e ATP-dipendente. Invece, in condizioni di stress metabolico in cui le proteine appena sintetizzate hanno difficoltà a maturare normalmente, si forma un legame stabile tra esse e una hsp 70, che funge da scorta.

L'idea che membri della famiglia dell'hsp 70 partecipino agli stadi precoci della maturazione delle proteine confermava i risultati di studi compiuti su una differente famiglia di proteine da stress. Lavori pionieristici svolti da Costa Georgopoulos dell'Università dello Utah e da altri avevano indicato che mutazioni nei geni responsabili della sintesi di due proteine da stress affini, la groEL e la groES, rendono i batteri incapaci di provvedere alla crescita di piccoli virus dipendenti dalla loro «macchina» cellulare. In assenza di groEL o di groES funzionali, molte proteine virali non riescono ad assemblarsi adeguatamente.

Infine, proteine simili alle proteine da stress batteriche groEL e groES sono state trovate anche in cellule di lieviti e in cellule animali e vegetali. Denominate hsp 10 e hsp 60, sono state localizzate solo nei mitocondri e nei cloroplasti. Dati recenti fanno pensare che ulteriori forme siano probabilmente presenti in altri compartimenti cellulari.

Indagini biochimiche hanno dimostrato inequivocabilmente che l'hsp 10 e l'hsp 60 sono essenziali per l'avvolgimento e l'assemblaggio delle proteine. La molecola dell'hsp 60 è costituita da due anelli sovrapposti, ciascuno contenente sette membri. Questa struttura

L'avvolgimento delle proteine avviene spontaneamente a causa dei vincoli termodinamici imposti dalla sequenza di amminoacidi idrofili e idrofobi nella molecola proteica. Nel corso del processo possono occasionalmente verificarsi errori, ma, di norma, le proteine si avvolgono in configurazioni funzionali sotto il profilo biologico (autoassemblaggio). Sembra che le proteine da stress contribuiscano ad assicurare un avvolgimento rapido ed estremamente fedele delle proteine cellulari.

sembra servire da «banco di lavoro» su cui le proteine non avvolte formano legami e acquisiscono la struttura tridimensionale definitiva. Secondo le concezioni correnti, l'avvolgimento è un processo estremamente dinamico e comporta una serie di eventi in cui si formano e si rompono legami. Ogni evento richiede energia, fornita dalla scissione enzimatica dell'ATP, e la partecipazione delle piccole molecole dell'hsp 10. Attraverso una serie di fasi di formazione e di scissione di legami, la proteina subisce cambiamenti conformazionali che la portano infine a raggiungere una forma stabile, opportunamente avvolta.

Si pensa che tanto la famiglia dell'hsp 60 quanto quella dell'hsp 70 operino insieme allo scopo di facilitare la maturazione delle proteine. Quando un nuovo polipeptide emerge da un ribosoma, è probabile che, all'interno di un organello o nel citoplasma, esso si leghi a una forma di hsp 70. Questa interazione può impedire che la catena polipeptidica in fase di allungamento si avvolga prematuramente. Una volta completata la sintesi, il nuovo polipeptide ancora legato alla sua scorta, l'hsp 70, sarebbe trasferito su una forma di hsp 60, su cui inizierebbero l'avvolgimento e l'assemblaggio con altre componenti proteiche.

Queste nuove osservazioni sulle proprietà dell'hsp 70 e dell'hsp 60 hanno costretto a riconsiderare i precedenti modelli di avvolgimento delle proteine. Ricerche effettuate negli anni cinquanta e sessanta avevano permesso di stabilire che una proteina denaturata può riavvolgersi spontaneamente dopo la rimozione dell'agente denaturante. Questi risultati portarono al concetto di autoassemblaggio delle proteine per il quale, nel 1972, fu conferito il premio Nobel per la chimica a Christian B. Anfinsen. Secondo il modello di Anfinsen, il processo di avvolgimento è dettato unicamente dalla sequenza degli amminoacidi nei polipeptidi. Gli amminoacidi idrofobi (non solubili in acqua) si posizionerebbero all'interno della molecola in fase di avvolgimento, mentre gli amminoacidi idrofili (solubili in acqua) si porterebbero alla superficie della proteina per poter essere sicuramente esposti all'ambiente acquoso della cellula. Pertanto l'avvolgimento procederebbe sotto il controllo diretto di restrizioni termodinamiche.

L'autoassemblaggio è tuttora considerato la forza principale che spinge le proteine ad assumere la loro conformazione definitiva. Oggi, tuttavia, molti ricercatori suppongono che per l'avvolgimento delle proteine sia necessaria la partecipazione di altre componenti cellulari, tra cui membri delle famiglie di proteine da stress hsp 60 e hsp 70.

Di conseguenza R. John Ellis dell'Università di Warwick e altri ricercatori hanno cominciato a chiamare la hsp 60, la hsp 70 e altre proteine da stress «chaperon molecolari» o chaperonine. Esse, pur non recando informazioni sull'avvolgimento o sull'assemblaggio delle proteine, assicurano che questi processi si svolgano con celerità e in modo estremamente fedele. La loro funzione è di accelerare l'autoassemblaggio riducendo la possibilità che una proteina in fase di maturazione imbocchi una via di avvolgimento errata.

Avendo stabilito che alcune proteine da stress fungono da chaperon molecolari in cellule sane, o comunque non soggette a stress, si è cercato di determinare perché in condizioni di stress quelle proteine siano espresse a livelli più elevati. Un indizio è dato dalle condizioni che ne incrementano l'espressione. All'interno delle cellule, temperature sufficienti ad attivare la risposta allo stress possono finire per denaturare alcune proteine. Le proteine denaturate dal calore, come quelle appena sintetizzate e non ancora avvolte, rappresenterebbero quindi bersagli a cui le proteine hsp 70 e hsp 60 possono legarsi. Col passare del tempo, via via che un numero maggiore di proteine denaturate dal calore si lega alle hsp 60 e hsp 70, i livelli di chaperon molecolari disponibili scendono e cominciano a limitare la capacità della cellula di produrre nuove proteine. La cellula avverte in qualche modo questa limitazione e risponde con un aumento della sintesi di nuove proteine da stress che fungono da chaperon molecolari.

Si ipotizza anche che un aumento nell'espressione delle proteine da stress possa essere indispensabile alle cellule per ristabilirsi dopo un danno metabolico. Se il calore, o altra causa di danno metabolico, denatura irreversibilmente molte proteine cellulari, la cellula deve correre ai ripari sostituendole. Così, l'aumento del livello di quelle proteine da stress che fungono da chaperon molecolari contribuisce a facilitare la sintesi e l'assemblaggio di nuove proteine. Inoltre livelli più elevati di proteine da stress possono impedire la denaturazione termica di altre proteine cellulari.

La riparazione e la sintesi di proteine sono in sé attività di vitale importanza. Tuttavia le proteine da stress hanno anche un ruolo peculiare nella regolazione di altri sistemi di proteine e di risposte cellulari. Un'altra famiglia di proteine da stress, il cui membro più conosciuto è la hsp 90, è particolarmente degna di nota sotto questo aspetto.

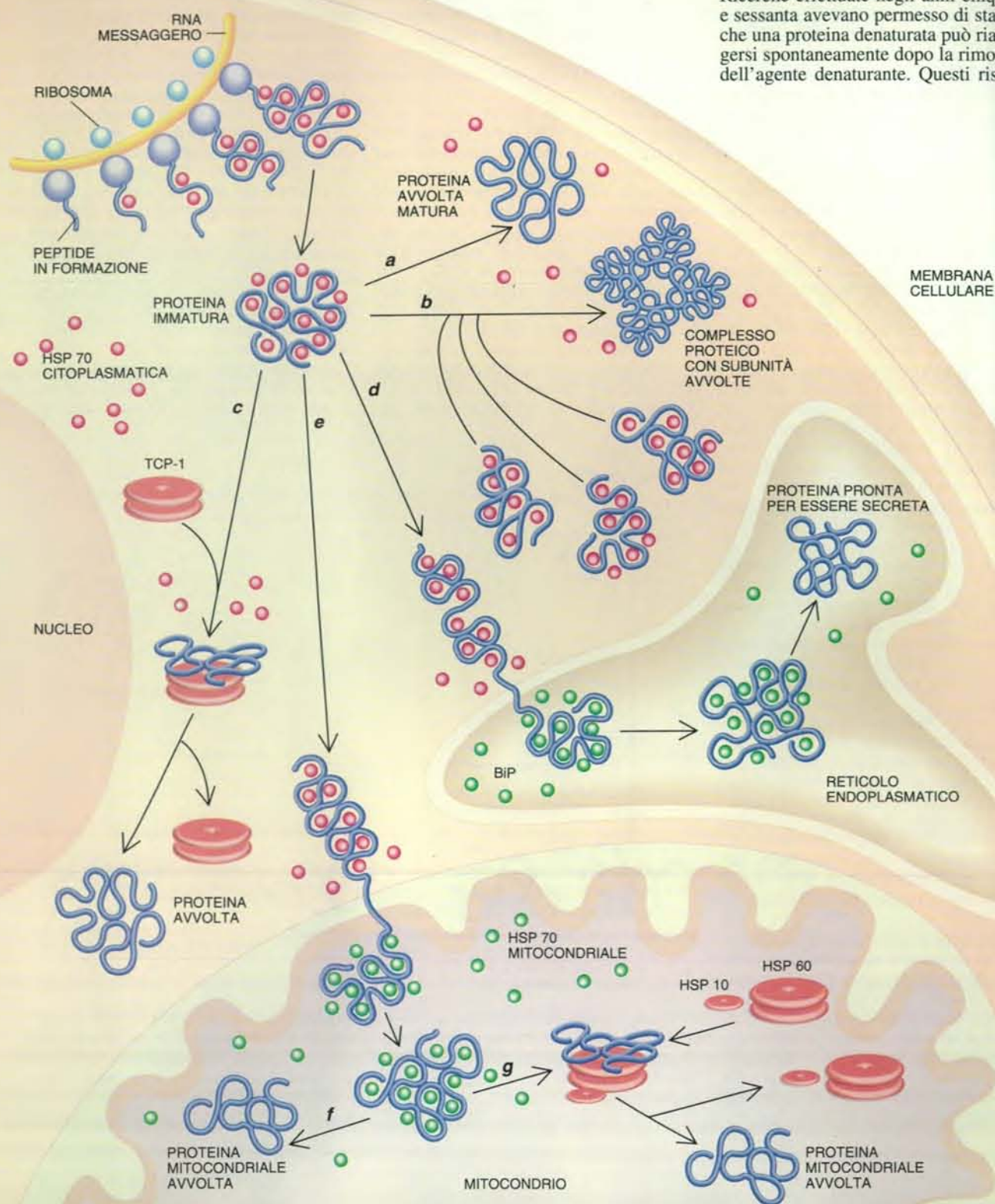
All'inizio l'interesse per l'hsp 90 è stato alimentato da resoconti che associavano questa proteina ad alcuni virus oncogeni. Alla fine degli anni settanta e all'inizio degli anni ottanta, gli oncologi hanno rivolto una considerevole attenzione al meccanismo con cui certi virus infettano le cellule facendole diventare neoplastiche. Nel caso del virus del sarcoma di Rous, è stato individuato un gene virale responsabile della comparsa di proprietà tumorali. L'enzima da esso prodotto, il pp60src, agisce su altre proteine che probabilmente regolano lo sviluppo cellulare. Da tre gruppi di ricerca è giunta indipendentemente la notizia che, dopo essere stato sintetizzato nel citoplasma, il pp60src si associa rapidamente a due proteine, l'una chiamata p50 e l'altra hsp 90.

Quando il pp60src si trova nel citoplasma ed è legato alle sue due molecole scorta, risulta inattivo come enzima. Non appena il terzetto molecolare si sposta e raggiunge la membrana cellulare, però, l'hsp 90 e la p50 si staccano e permettono al pp60src di depositarsi nella membrana stessa diventando attivo. Analoghe interazioni sono state scoperte tra l'hsp 90, la p50 ed enzimi oncogeni codificati da parecchi altri virus che provocano tumori. Sembra che questi enzimi virali, quando sono legati all'hsp 90 e alla p50, siano incapaci di agire sui bersagli cellulari necessari perché si sviluppi lo stato tumorale.

Ulteriori studi hanno anche stabilito un collegamento tra l'hsp 90 e un'altra importante classe di molecole presenti nelle cellule dei mammiferi, i recettori degli ormoni steroidei. Negli animali questi ormoni mediano processi biologici di vitale importanza. Per esempio, i glicocorticoidi contribuiscono a sopprimere l'infiammazione, mentre altri ormoni steroidei svolgono funzioni importanti nel differenziamento e nello sviluppo sessuale. Quando un recettore degli ormoni steroidei si lega al proprio ormone specifico, diventa capace di interagire con il DNA e di attivare o reprimere l'espressione di certi geni.

Un interrogativo cruciale era in che modo i recettori vengano mantenuti inattivi all'interno di una cellula. La risposta è emersa in modo chiaro con la caratterizzazione delle forme attiva e inattiva del recettore per il progesterone. In assenza dell'ormone, il recettore si associa a parecchie altre proteine cellulari, tra cui l'hsp 90, che lo mantengono in uno stato inattivo. Dopo il legame con il progesterone, il recettore si libera dall'hsp 90 e va incontro a una serie di cambiamenti che gli permettono di legarsi con il DNA. Come nel caso degli enzimi virali, sembra che l'hsp 90 regoli l'attività biologica dei recettori degli ormoni steroidei.

Si comincia solo oggi a valutare le applicazioni pratiche della risposta allo stress, che sembrano particolarmente promettenti in campo medico. Quando un individuo è colpito da un infarto miocardico o cerebrale, l'irrorazione sanguigna del cuore o del cervello è temporaneamente compromessa: questa condizione viene definita ischemia. In mancanza di ossigeno, l'organo colpito non può mantenere i propri livelli normali di ATP, il che fa venir meno processi metabolici essenziali. Quando il flusso sanguigno si ristabilisce, l'organo ischemico viene rapidamente riossigenato, ma anche questo può essere dannoso. Spesso la rapida riesposizione all'ossigeno genera specie molecolari estremamente reattive, i radicali liberi, che possono compiere ulteriori danni.

In studi su animali si sono osservate

Parecchi meccanismi per l'avvolgimento e la distribuzione delle proteine all'interno delle cellule sono sotto il controllo di proteine da stress. In molti casi sembra che differenti proteine di questo tipo operino in tandem. Nel citoplasma una forma di hsp 70 si lega alle proteine nel momento in cui queste vengono prodotte dai ribosomi per impedire un loro avvolgimento prematuro. L'hsp 70 può poi dissociarsi dalla proteina e lasciare che essa si avvolga nella sua forma funzionale (*a*) o che si associ con altre proteine formando complessi multimerici più grandi (*b*). In alcuni casi le proteine vengono trasferite dall'hsp 70 a un'altra proteina da stress, la TCP-1, prima che avvengano l'avvolgimento e l'assemblaggio finali (*c*). Se la proteina è destinata alla secrezione, può essere trasportata nel reticolo endoplasmatico e trasferita alla BiP o a un'altra proteina da stress affine che ne dirige l'avvolgimento definitivo (*d*). Altre proteine vengono trasferite ai mitocondri o ad altri organelli (*e*). All'interno dei mitocondri un'altra forma specializzata di hsp 70 assiste talvolta la proteina nell'avvolgimento definitivo (*f*), ma in molti casi quest'ultima viene affidata a un complesso costituito da hsp 60 e hsp 10 (*g*). Sembra che la molecola di hsp 60 funga da forma attorno alla quale la proteina mitocondriale può avvolgersi.

risposte allo stress indotte sia nel cuore sia nel cervello dopo brevi episodi di ischemia e di riperfusione. Sembra che l'entità di queste risposte sia da correlare direttamente con la gravità relativa del danno. Pertanto si comincia a prendere in considerazione l'utilità delle variazioni dei livelli delle proteine da stress quali indici di lesioni a tessuti e organi.

Sembra che le cellule che producono livelli elevati di proteine da stress siano in grado di resistere meglio al danno ischemico rispetto alle altre cellule. Di conseguenza, elevando i livelli di queste proteine, per esempio con mezzi farmacologici, si può offrire un'ulteriore protezione ai tessuti e agli organi lesi. Una simile impostazione terapeutica potrebbe ridurre il danno da ischemia ai tessuti, che sopraggiunge nel corso di un intervento chirurgico, oppure contribuire a proteggere gli organi utilizzati per i trapianti, che spesso soffrono di danni da ischemia e riperfusione.

Una novità estremamente interessante riguarda il ruolo della risposta allo stress nell'immunità e nelle malattie infettive. La tubercolosi, la malaria, la lebbra, la schistosomiasi e altre malattie che colpiscono ogni anno milioni di individui sono conseguenza di un'infezione da parte di batteri o di altri microrganismi parassiti. Si è scoperto che le proteine da stress prodotte da questi organismi sono spesso i principali antigeni, o proteine bersaglio, che il sistema immunitario utilizza per riconoscere e distruggere gli invasori. Può darsi che il sistema immunitario umano sia costantemente alla caccia di forme estranee di proteine da stress. Le proteine da stress di vari agenti patogeni, prodotte in laboratorio con le tecniche del DNA ricombinante, potrebbero dunque potenzialmente essere impiegate come vaccini per prevenire le infezioni microbiche. Inoltre, per la loro azione altamente immunogenica, le proteine da stress microbiche potrebbero essere utilizzate come adiuvanti: legate alle proteine virali, potrebbero potenziare le risposte immunitarie contro le infezioni virali.

Gli immunologi hanno anche scoperto una possibile connessione tra proteine da stress e malattie autoimmuni. La maggior parte di queste patologie insorge quando il sistema immunitario reagisce ad antigeni presenti nei tessuti sani. In alcune di esse, tra cui l'artrite reumatoide, la spondilite anchilosante e il lupus eritematoso sistemico, si osservano talvolta anticorpi contro le proteine da stress del paziente. Se queste osservazioni saranno confermate per un gran numero di soggetti, potrebbero risultare utili nella diagnosi e forse anche nella cura delle malattie autoimmuni.

Dato che le proteine da stress dei microrganismi sono molto simili alle proteine da stress umane, può darsi che il sistema immunitario umano sia costantemente costretto a discernere piccole differenze tra le proteine da stress dell'organismo e quelle dei microrganismi invasori. La possibilità che le proteine da stress si situino in posizione privilegiata all'interfaccia tra tolleranza a un organismo infettivo e autoimmunità è un'idea interessante, che continua a suscitare dibattiti tra i ricercatori.

La presenza di anticorpi diretti contro proteine da stress microbiche può risultare utile nella diagnostica. Per esempio, il batterio *Chlamydia trachomatis* provoca numerose malattie, fra cui il tracoma, che è probabilmente, in tutto il mondo, la causa più importante di cecità prevenibile, e un'infiammazione pelvica che è causa rilevante di sterilità nella donna. L'infezione da *Chlamydia* innesca di solito la produzione di anticorpi contro gli antigeni batterici, alcuni dei quali sono proteine da stress. Spesso questa risposta immunitaria è efficace e finisce per eliminare l'agente patogeno; in alcuni individui, però, particolarmente in quelli che hanno subito infezioni ripetute o croniche da *Chlamydia*, essa risulta eccessivamente aggressiva e provoca lesioni e danni permanenti nei tessuti circostanti.

Richard S. Stephens e collaboratori dell'Università della California a San Francisco hanno osservato che oltre il 30 per cento delle donne colpite da infiammazioni alla pelvi e oltre l'80 per cento di quelle che hanno avuto gravidanze ectopiche possiedono livelli eccezionalmente alti di anticorpi diretti contro la proteina da stress groEL di *Chlamydia*. La misurazione di questi anticorpi può risultare utile per identificare donne ad alto rischio di gravidanze ectopiche o di sterilità.

Il legame tra proteine da stress, risposta immunitaria e malattie autoimmuni diventa ancora più interessante alla luce di altre recenti scoperte. Alcuni membri della famiglia hsp 70 sono molto simili per struttura e funzione agli antigeni di istocompatibilità. Queste ultime proteine prendono parte ai primissimi stadi della risposta immunitaria, presentando gli antigeni estranei alle cellule del sistema immunitario.

Ci si è chiesti in che modo un singolo antigene di istocompatibilità possa legarsi a una schiera molto differenziata di peptidi antigenici. Di recente Don C. Wiley e collaboratori dell'Università di Harvard hanno contribuito a risolvere questo problema determinando la struttura tridimensionale delle proteine di istocompatibilità di classe I. Essi hanno trovato che un solco o incavo nella molecola delle suddette proteine può legarsi a differenti peptidi antigenici. Nello stesso tempo James E. Rothman, allora alla Princeton University, ha riferito che i membri della famiglia di proteine da stress hsp 70 sono anch'essi in grado di legarsi a brevi peptidi. Questa proprietà è conforme con il ruolo che le hsp 70 svolgono nel legarsi ad alcune parti di catene polipeptidiche non avvolte o appena sintetizzate.

Modelli al calcolatore hanno rivelato che la proteina hsp 70 ha probabilmente un sito di legame per i peptidi analogo a quello delle proteine di istocompatibilità di classe I. L'evidente rassomiglianza tra le due classi di proteine appare ancora più interessante perché parecchi dei geni che codificano per le hsp 70 sono localizzati molto vicino ai geni per le proteine di istocompatibilità. Considerate nel loro complesso, tutte queste osservazioni rafforzano l'ipotesi che le proteine da stress siano parte integrante del sistema immunitario.

La capacità di manipolare la risposta allo stress può dimostrarsi importante anche nell'elaborare nuovi metodi terapeutici contro il cancro. Spesso i tumori sembrano avere una sensibilità termica superiore a quella dei tessuti normali. Elevare la temperatura dei tessuti per debellare i tumori è un'idea ancora allo stadio sperimentale, ma in esperimenti preliminari l'uso dell'ipertermia localizzata, sola o associata alla radioterapia o ad altre terapie convenzionali, ha portato alla regressione di certi tipi di tumori.

La risposta allo stress però non è necessariamente un'alleata del medico nella cura del cancro: può anche essere un ostacolo. Dato che le proteine da stress danno alle cellule una protezione in più, le terapie antitumorali che inducono una risposta allo stress possono rendere il tumore più resistente a trattamenti successivi. Tuttavia non è detto che non si scopra qualche metodo per inibire la capacità di un tumore di scatenare una risposta allo stress, privandolo così delle sue difese contro una particolare terapia.

Oggi si sta anche cominciando a considerare la possibilità di utilizzare la risposta allo stress in tossicologia. Un cambiamento nel livello delle proteine da stress, particolarmente di quelle prodotte soltanto da cellule traumatizzate, può risultare utile per stabilire la tossicità di farmaci, cosmetici, additivi alimentari e altri prodotti. La ricerca in questo senso è solo a uno stadio preliminare di sviluppo, ma parecchie strategie di applicazione stanno già dando segni di successo.

Ricorrendo alla tecnologia del DNA ricombinante, sono state ottenute in coltura genealogie di cellule «indicatrici di stress» che potrebbero essere utilizzate nella valutazione dei rischi biologici. In queste cellule le sequenze di DNA che controllano l'attività dei geni che codificano per le proteine da stress sono legate a un gene indicatore che codifica per un enzima, come la beta-galattosidasi. Quando queste cellule subiscono uno stress metabolico e producono una maggiore quantità di proteine da stress, sintetizzano anche l'enzima indicatore, che può essere facilmente identificato con vari saggi. La quantità di beta-galattosidasi espressa in una cellula può essere misurata aggiungendo un substrato chimico. Se l'enzima indicatore è presente, la cellula diventa blu e l'intensità del colore è direttamente proporzionale alla concentrazione dell'enzima al suo interno. Servendosi di queste cellule indicatrici, si può facilmente determinare l'entità della risposta allo stress indotta da agenti o trattamenti chimici. Se risultassero affidabili, saggi di questo tipo potrebbero un giorno ridurre o perfino sostituire l'impiego di animali nei test tossicologici.

La stessa tecnica potrebbe anche essere estesa al controllo dei rischi da inquinanti ambientali, molti dei quali scatenano risposte allo stress. A questo scopo si sono cominciati a produrre organismi transgenici indicatori di stress. Eve G. Stringham ed E. Peter M. Candido dell'Università della British Columbia, in collaborazione con la Stressgen Biotechnologies di Victoria, hanno prodotto vermi transgenici nei quali un gene indicatore per la beta-galattosidasi è sotto il controllo del promotore di una proteina da shock termico. Quando questi organismi sono esposti a vari agenti inquinanti, esprimono l'enzima indicatore e diventano blu-verdi. Nel laboratorio di Candido si sta cercando di determinare se questi organismi indicatori di stress possano essere utili nel monitoraggio di una vasta gamma di agenti inquinanti.

Voellmy e Nicole Bournias-Vardiabasis, allora al City of Hope National Medical Center di Duarte, in California, hanno utilizzato un procedimento analogo per produrre moscerini della frutta transgenici indicatori di stress. Questi moscerini diventano blu quando sono esposti ad agenti teratogeni, ossia sostanze che provocano uno sviluppo fetale anomalo. È significativo il fatto che questi insetti siano sensibili a molti agenti teratogeni che notoriamente causano difetti congeniti nell'uomo. Sembra che la strada sia ormai aperta per la produzione di altri organismi indicatori di stress che potrebbero risultare utili nelle analisi tossicologiche e ambientali.

Oltre 30 anni fa, le risposte allo stress e allo shock termico nel moscerino della frutta sembravano semplici curiosità molecolari; oggi esse sono al centro di un campo di ricerca attivo e vitale. Gli studi sulla struttura e sulla funzione delle proteine da stress hanno portato a nuove conoscenze su processi cellulari essenziali, comprese le vie lungo cui si svolge la maturazione delle proteine. Si sta anche imparando a utilizzare le conoscenze acquisite sulla risposta allo stress per risolvere problemi in campo medico e ambientale. Ho la sensazione che si cominci solo ora a rendersi conto di tutte le implicazioni di questo antichissimo meccanismo di risposta, grazie al quale le cellule riescono a far fronte allo stress.

Le risposte agli ormoni steroidei sono in parte controllate dalla hsp 90. Questa proteina da stress contribuisce a mantenere inattivi i recettori per i suddetti ormoni. Quando gli ormoni steroidei sono presenti, si legano al recettore e l'hsp 90 viene liberata. A questo punto il complesso ormone-recettore può interagire con il DNA e avviare l'espressione dei geni che codificano per certe proteine.

Alcune condizioni che inducono l'espressione di proteine da stress

FATTORI AMBIENTALI
- SHOCK TERMICO
- METALLI PESANTI DI TRANSIZIONE
- INIBITORI DEL METABOLISMO ENERGETICO
- ANALOGHI DEGLI AMMINOACIDI
- AGENTI CHEMIOTERAPICI

STATI PATOLOGICI
- INFEZIONI VIRALI
- FEBBRE
- INFIAMMAZIONE
- ISCHEMIA
- IPERTROFIA
- LESIONI DA OSSIDANTI
- NEOPLASIE

FATTORI CELLULARI NORMALI
- CICLO DI DIVISIONE CELLULARE
- FATTORI DI CRESCITA
- SVILUPPO E DIFFERENZIAMENTO

BIBLIOGRAFIA

ASHBURNER M. e BONNER J. J., *The Induction of Gene Activity in* Drosophila *by Heat Shock* in «Cell», 17, n. 2, giugno 1979.

MORIMOTO RICHARD I., TISSIÈRES ALFRED e GEORGOPOULOS COSTA, *Stress Proteins in Biology and Medicine*, Cold Spring Harbor Laboratory Press, 1990.

ELLIS R. JOHN e VAN DER VIES S. M., *Molecular Chaperones* in «Annual Reviews of Biochemistry», 60, pp. 321-347, 1991.

LANGER THOMAS, CHI LU, ECHOLS HARRISON, FLANAGAN JOHN, HAYER MANAJIT K. e HARTL F. ULRICH, *Successive Action of DnaK, DnaJ and GroEL along the Pathway of Chaperone-Mediated Protein Folding* in «Nature», 356, n. 6371, 23 aprile 1992.

WELCH WILLIAM J., *Mammalian Stress Response: Cell Physiology, Structure-/Function of Stress Proteins, and Implications for Medicine and Disease* in «Physiological Reviews», 72, pp. 1063-1081, ottobre 1992.

Il valore informativo delle carte geografiche

La pretesa oggettività delle rappresentazioni cartografiche moderne ne maschera le intrinseche distorsioni

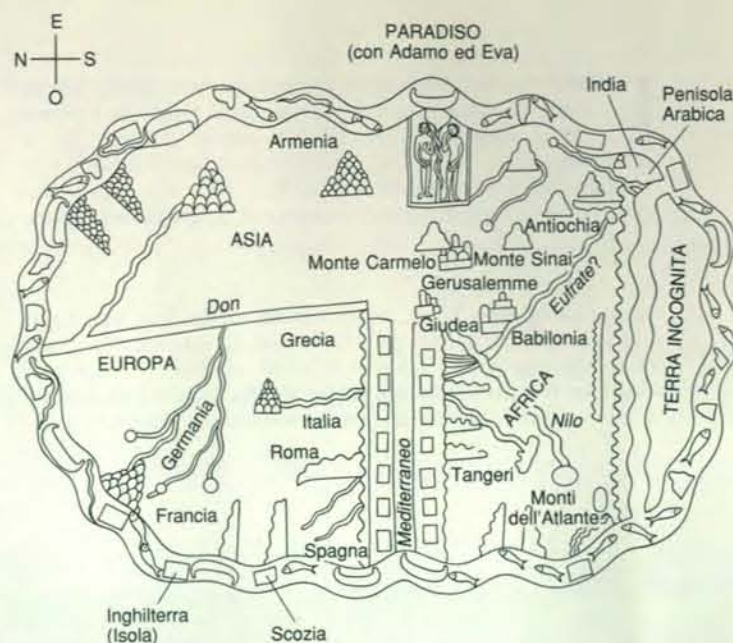
di Denis Wood

L'obiettività delle moderne rappresentazioni cartografiche del mondo è data per scontata a tal punto da fungere quasi da metafora per le altre scienze, e magari per la stessa obiettività scientifica. La visione tradizionale della storia della cartografia occidentale tende a rafforzare questa pretesa di obiettività. La storia ci racconta di un graduale progres-

Questa carta della Terra è stata realizzata da Tom Van Sant, «setacciando» 37 milioni di pixel inviati da satellite. Van Sant ha manipolato i pixel in modo da eliminare le nubi e porre in risalto certi aspetti naturali. La fotografia della NASA mostra la Terra vista dallo spazio, con il suo abituale manto di nubi.

so a partire dalle grezze visioni medievali del mondo fino appunto alle rappresentazioni che si allineano agli standard attuali di precisione. In realtà, tutte le carte geografiche incorporano assunti e convenzioni propri della società e degli individui che le hanno prodotte. Queste influenze appaiono ovvie nelle mappe antiche, ma tendono di solito a passare inosservate nelle carte dei tempi moderni. In realtà, solo la consapevolezza delle omissioni e distorsioni soggettive inerenti alle rappresentazioni cartografiche può permettere a chi ne fa uso di ricavare un senso pieno dall'informazione che esse contengono.

La storia della cartografia viene fatta comunemente risalire agli egizi e ai babilonesi. Passando attraverso i contributi degli antichi greci e romani, si riconoscono grandi meriti all'apporto degli arabi durante il Medioevo. Si è ritenuto a lungo che la cartografia avesse raggiunto il suo punto più basso nell'Europa medievale. Dal XIV secolo in avanti si ebbero lenti e costanti progressi, fino ad arrivare alle carte dei nostri giorni, che si avvalgono dell'uso di ottiche sofisticate, di immagini da satellite e di elaborazioni digitali.

Data la lunga tradizione della rappresentazione cartografica, potrebbe apparire sorprendente l'affermazione che ben pochi oggetti dei tempi remoti possono essere identificati con certezza come mappe. Sono giunti sino a noi numerosi frammenti di testi che testimoniano come in Grecia si realizzassero mappe, ma non siamo in possesso di alcuna di esse. Fatta eccezione per le copie medievali di itinerari romani - di cui le più famose sono *La carta di Madaba*, un esteso mosaico pavimentale del VI secolo rinvenuto in un'antica basilica della Giordania e la

Le antiche rappresentazioni cartografiche riflettono con chiarezza le concezioni dei loro autori. La mappa di Beato (*nella pagina a fronte, con spiegazioni in alto*) ritraeva la visione cristiana del mondo nell'VIII secolo e pertanto include il Paradiso terrestre, collocandolo a est. Nel XIV secolo, le mappe tolemaiche (*in alto a destra*) segnarono un cambiamento verso una rappresentazione del mondo che meglio sarebbe servita agli interessi del nascente commercio su scala mondiale centrato sull'Europa. Le carte del XIX secolo, come quella tratta dall'atlante di Cary, riprodotta al centro, presupponevano interessi commerciali e politici. Da parte loro, i cartografi orientali ponevano la Cina al centro del mondo, come nell'esempio in basso raffigurante una carta coreana del secolo scorso.

Tabula Peutingeriana ritrovata solo nel 1507 - non si conoscono rappresentazioni romane del mondo, nonostante le elaborate istruzioni per la loro realizzazione che si ritrovano nella *Geografia* di Tolomeo. Del resto, anche nel Medioevo non sono molte le mappe che possano servire da riferimento per valutare il livello a cui era giunta la cartografia in quel periodo.

Queste mappe del mondo medievali, o *mappaemundi*, si presentano in molte varietà. La mappa di Beato, disegnata originariamente nel 776 per illustrare il *Commento all'Apocalisse di San Giovanni* dell'abate Beato di Libana e poi andata perduta, è esemplare. Fortunatamente, di questa mappa sono giunte a noi alcune copie del X secolo, o anche posteriori. La mappa di Beato riprodotta nella pagina a fronte è rettangolare ma, come le altre *mappaemundi*, è orientata in modo che l'est, dove è collocato il Paradiso Terrestre, sia in alto. I tre continenti, popolati dai figli di Noè, sono disposti secondo la tradizione: l'Europa in basso a sinistra; l'Africa in basso a destra e l'Asia sopra di essi. La mappa di Beato considera anche un quarto continente, la *Terra incognita*, la cui esistenza era supposta in base al testo evangelico (agli apostoli era stato dato mandato di predicare il Vangelo «ai quattro angoli della Terra»).

In primo luogo e sopra ogni altra considerazione, la mappa di Beato fornisce una visione della Terra come teatro della storia cristiana del mondo; la precisione geografica è di interesse del tutto secondario. Rispetto a un atlante di Rand McNally, la mappa di Beato appare affatto erronea, ma non avrebbe senso asserire che le carte geografiche più recenti racchiudano un senso «più vero» del mondo: data la loro funzione religiosa, le mappe di Beato sono del tutto accurate.

La riscoperta dei testi di Tolomeo sul modo di realizzare mappe, avvenuta con l'inizio delle Crociate, condusse nel XIV secolo a uno stile cartografico di aspetto più simile a quello moderno. I cartografi, seguendo le istruzioni di Tolomeo e avvalendosi dei suoi dati, produssero mappe che presentavano il nord in alto e mostravano località fissate in un reticolo che permetteva di ricavarne latitudine e longitudine. Per quanto presentino una superficiale somiglianza con le carte dei nostri giorni, le mappe tolemaiche rimanevano profondamente segnate dalle convenzioni delle *mappaemundi* medievali. Continuavano a mostrare una *Terra incognita* a sud, facevano grande uso di figure decorative (come quelle che rappresentavano i dodici venti) e impiegavano convenzioni cromatiche tradizionali (per esempio, il Mar Rosso si poteva dipingere in rosso). Comunque queste mappe segnarono un cambiamento rispetto all'impegno di interpretare il mondo attraverso l'ottica biblica, in favore di finalità più pratiche, che meglio servivano al nascere di un commercio mondiale centrato sull'Europa.

Le carte del XVIII e del XIX secolo presupponevano gli interessi commerciali e politici degli Stati-nazione europei. Nei sempre più diffusi atlanti del XIX secolo appare un mondo eurocentrico. Le frontiere, i simboli, le illustrazioni e le note che si vedono su queste carte esprimono graficamente gli interessi politici, commerciali e scientifici degli stati europei; si mettevano in particolare evidenza i possedimenti coloniali. Queste carte, realizzate sulla base della tradizione tolemaica, stabiliscono un nuovo insieme di convenzioni. Il nord è in alto; gli zero gradi di longitudine corrispondono al meridiano di

Greenwich, in Inghilterra, e le carte sono centrate sull'Europa occidentale, il Nord America e l'Atlantico settentrionale. La configurazione risultante è divenuta così familiare da farci perdere la cognizione di quanto in realtà sia arbitraria.

Con il trasformarsi della cartografia in un'arte scientifica, per l'uomo comune diviene sempre più difficile rendersi conto del fatto che le carte geografiche sono finestre, oltre che sul mondo che rappresentano, anche sulla società che le ha prodotte. Nell'Occidente, l'epistemologia positivista dominante e la diffusa credenza in un continuo progresso materiale hanno fatto sì che gli storici e i profani fossero portati a considerare primitive le carte non-occidentali e a denigrare le antiche mappe come prodotti di una mentalità barbarica e superata. Per estensione logica, si suppone che le carte del tempo presente siano le migliori, le più precise e oggettive mai prodotte. Questo modo di vedere è stato notevolmente rafforzato dalla corrispondenza perfetta, fino alla possibilità di confusione, fra le carte e le immagini della Terra riprese da satelliti artificiali.

Si consideri *GeoSphere*, una carta realizzata da Tom Van Sant, che lavora al GeoSphere Project a Santa Monica, in California, con l'assistenza tecnica di Lloyd Van Warren, del Jet Propulsion Laboratory di Pasadena. Con essa viene meno la linea di demarcazione fra la carta e l'immagine da satellite. Van Sant e Van Warren hanno realizzato la carta «setacciando» milioni di pixel trasmessi dai satelliti *Tiros-N*, gestiti dalla National Oceanic and Atmospheric Administration. Essi hanno scartato le immagini in cui la presenza di nubi oscurava la superficie terrestre, lasciando apparire contorni continentali definiti da un occhio elettronico imparziale.

Ma *GeoSphere* incorpora un suo contenuto ideologico in ogni bit, esattamente come facevano le *mappaemundi*, le mappe tolemaiche e gli atlanti del XIX secolo. Come molti suoi predecessori, Van Sant ha scelto di far decorrere l'equatore a metà della sua carta, di porre l'Atlantico al centro e di orientarlo con il nord in alto. Inoltre Van Sant riconosce apertamente di avere filtrato e modificato i dati da satellite in maniera deliberatamente soggettiva.

La stessa esclusione delle nubi è un'omissione di una caratteristica dominante della Terra, così come si vede dallo spazio. Nelle zone per cui non erano disponibili immagini prive di nubi, i cartografi hanno provveduto a sottrarle artificialmente, pixel dopo pixel. Per latitudini basse e intermedie, hanno selezionato quelle immagini che mostravano in modo più pronunciato la vegetazione estiva; per elevate latitudini e altitudini, hanno scelto immagini che esaltavano l'innevamento invernale. I sistemi fluviali sono stati ispessiti, per essere resi più visibili, e sono stati applicati falsi colori, paradossalmente, per rendere più reale l'idea della copertura vegetale. Queste decisioni seguono un criterio di utilità, in quanto fanno risaltare certi aspetti della Terra, rendendo quindi la carta di Van Sant più utile e di facile lettura di quanto sarebbe altrimenti. Ma si dovrebbe tenere a mente che l'assenza di nubi, l'estensione della copertura vegetale, la visibilità dei fiumi e tutti i colori impiegati sono espressioni della visione personale del cartografo, e non attributi intrinseci della Terra stessa.

L'ombreggiatura e la colorazione della carta di Van Sant enfatizzano gli aspetti naturali della Terra (vegetazione, deserto, neve e ghiaccio), omettendo i segni della presenza umana. W. T. Sullivan, dell'Università di Washington, ha realizzato una carta pressoché antitetica che dà il massimo risalto all'impatto antropico. La carta di Sullivan, che è stata pubblicata dall'Hansen Planetarium, si avvale anch'essa di immagini «depurate» dalle nubi trasmesse dai satelliti *Tiros-N*, ma in questo caso le immagini sono state riprese sul lato in ombra della Terra. Semplicemente passando a una prospettiva notturna, Sullivan ha eliminato i contorni continentali, lasciando solo le luci delle città e i fuochi a segnare i siti di attività umana.

Una terza carta, pure derivata dai dati da satellite, presenta un'altra visione del mondo, scientificamente precisa ma completamente differente. Realizzata da William F. Haxby mentre si trovava al Lamont Doherty Geological Observatory della Columbia University, la carta mostra le anomalie del campo gravitazionale terrestre del fondo oceanico, registrate per mezzo di rilevamenti radar altimetrici della superficie marina fatti dal satellite *Seasat*. Haxby ignora totalmente la superficie fisica della Terra. La carta di anomalia gravitazionale sembra mostrare la topografia del fondo oceanico. Si tratta di un'illusione, rafforzata dal modo in cui Haxby ha illuminato le dorsali e le valli, a luce radente. Questo tipo di rappresentazione rende i dati più comprensibili, ma reca in sé la possibilità di fuorviare un osservatore distratto.

La carta di Haxby proclama apertamente il proprio carattere soggettivo: le anomalie gravitazionali del fondo oceanico sono rese con vividi falsi colori, mentre i continenti risultano oscurati, privi di interesse. Per il fatto di escludere esplicitamente informazioni su certe parti della Terra, la carta di Haxby vede esaltata la propria qualità tematica. Si dovrebbe porre maggiore attenzione nell'interpretare carte che sembrano mostrare la Terra come essa apparirebbe a un osservatore esterno. Queste carte vengono spesso descritte come «ritratti» o «viste», e ciò rafforza quella falsa impressione.

Ogni carta in due dimensioni è soggettiva, dato che non si può evitare di distorcere in qualche modo le dimensioni e le forme di ciò che si rappresenta. I cartografi cercano di superare questo limite in modi diversi. La carta di Van Sant mantiene una relativa fedeltà di forma a spese della precisione nella rappresentazione delle misure; si tratta di una scelta appropriata ai fini che la carta si propone. Conservation International, viceversa, ha optato per una proiezione equivalente (che conserva le aree relative dei continenti, a costo di distorcerne la forma). Questa organizzazione, che si occupa della salvaguardia della foresta pluviale tropicale, ha scelto una proiezione che non esagera le dimensioni dell'Europa e del Nord America a spese di Africa tropicale, Asia e Sud America, così come avviene nelle proiezioni di uso più comune. Come la carta di Van Sant, anche questa adotta colorazioni particolari per dare risalto a certi aspetti naturali della Terra. Dati gli scopi che si propone, comunque, Conservation International usa ovviamente falsi colori che richiamano l'attenzione specificamente sulle regioni minacciate della foresta pluviale.

Ognuna delle carte moderne qui considerate vanta la propria speciale accuratezza, e purtuttavia i risultati potrebbero difficilmente essere più dissimili. Questa è la contraddizione intrinseca di ogni carta geografica: la pretesa di rappresentare oggettivamente un mondo che essa può mostrare solo in modo soggettivo. La pretesa viene avanzata al fine di guadagnare consenso per una visione del mondo la cui utilità risiede esattamente nella sua parzialità.

E dato che l'utilità delle carte deriva dalle loro distorsioni e parzialità, queste sono qualità che devono essere messe in risalto e vantate. Occorre che le carte geografiche siano esplicite riguardo alle scelte fatte tra i molti insiemi di dati disponibili e tra i vari modi di rappresentarli. Dovrebbe essere resa chiara la natura delle limitazioni e delle distorsioni di ogni carta, i vantaggi che essa presenta e ciò che essa tende a porre maggiormente in risalto. E chi fa uso delle carte dovrebbe essere meglio edotto sulle possibilità da esse offerte. Il futuro della cartografia risiede nelle possibilità di trascendere la dicotomia tra l'utilità del soggettivo e l'autorità dell'oggettivo. Si potrà così arrivare a produrre carte ancora più utili, in quanto più esplicite riguardo alla loro reale relazione con il mondo.

Ogni carta geografica moderna mostra distorsioni che servono a mettere in risalto le sue particolari finalità. Nella carta di William F. Haxby, che rappresenta le anomalie di gravità dei fondi oceanici, i continenti vengono considerati non significativi (*nella pagina a fronte*). La carta di Conservation International (*in alto a destra*) pone in risalto l'impatto antropico sugli ecosistemi e distorce la forma dei continenti per conservarne le aree. La carta notturna di W. T. Sullivan (*al centro*), realizzata con immagini da satellite riprese sul lato in ombra del nostro pianeta, mette in evidenza la presenza umana sulla Terra. Una carta fantasiosa, la «McArthur's Universal Corrective Map» (*in basso*), colloca il Polo Sud in alto, per mostrare il mondo dal punto di vista degli australiani.

BIBLIOGRAFIA

HARLEY J.B. e WOODWARD DAVID (a cura), *The History of Cartography*, vol. 1: *Cartography in Prehistoric, Ancient, and Medieval Europe and the Mediterranean*, University of Chicago Press, 1987.

HALL STEPHEN S., *Mapping the Next Millennium: The Discovery of New Geographies*, Random House, 1992.

WOOD DENIS, *The Power of Maps*, Guilford Press, 1992.

La neurobiologia della paura

Le recenti ricerche sui processi neurochimici che sono alla base delle diverse reazioni indotte dalla paura nelle scimmie potrebbero consentire di mettere a punto nuove terapie per curare gli stati d'ansia nell'uomo

di Ned H. Kalin

Nel corso della vita, la maggior parte delle persone acquisisce un repertorio di capacità che consentono di far fronte a tutta una varietà di situazioni che incutono paura. Un individuo può cercare di calmare l'ira di un insegnante o di un superiore, oppure può sfuggire urlando all'aggressore. Alcuni però si lasciano sopraffare da situazioni che altri considererebbero assai poco stressanti: la paura del ridicolo può farli tremare incontrollabilmente se vengono chiamati a parlare in pubblico, o il terrore degli estranei può indurli a nascondersi in casa, incapaci di andare al lavoro o di uscire a far spese. Perché alcune persone diventano vittime di paure eccessive?

Presso l'Università del Wisconsin a Madison, il mio collega Steven E. Shelton e io stiamo provando a dare una risposta a questo interrogativo attraverso l'identificazione di specifici processi cerebrali che regolano la paura e i comportamenti a essa associati. Nonostante la disponibilità di tecniche non invasive per la realizzazione di immagini cliniche, queste informazioni sono tuttora estremamente difficili da ottenere nell'uomo. Pertanto abbiamo rivolto la nostra attenzione a un altro primate, la scimmia reso (*Macaca mulatta*). Gli individui di questa specie attraversano molte delle fasi di sviluppo fisiologico e psicologico a cui va incontro l'uomo, sebbene le vivano in un arco di tempo più ristretto. Via via che chiariremo la natura e il funzionamento dei circuiti nervosi che regolano la paura nelle scimmie, dovremmo riuscire anche a individuare i processi cerebrali responsabili dell'ansia eccessiva nell'uomo e a escogitare nuove terapie per alleviare questa sindrome.

Può essere particolarmente vantaggioso intervenire precocemente: è stato infatti osservato che i bambini eccessivamente timorosi hanno un alto rischio di soffrire in seguito di disturbi emotivi. Jerome Kagan e colleghi della Harvard University hanno dimostrato, per esempio, che un bambino estremamente timido all'età di due anni ha maggiori probabilità rispetto a un bambino meno inibito di soffrire successivamente di ansia e depressione.

Ciò non significa che queste patologie siano inevitabili, ma è facile comprendere come chi soffre eccessivamente di paure possa andare incontro a una vita segnata da difficoltà emotive. Si consideri il caso di un bambino estremamente timoroso dei suoi coetanei che, per questa ragione, venga spesso canzonato a scuola; egli potrebbe cominciare a ritenersi una persona poco piacevole e, per reazione, a ritirarsi ulteriormente in se stesso. Col passar del tempo, potrebbe restare intrappolato in un circolo vizioso che lo condurrebbe all'isolamento, alla scarsa autostima, a cattivi risultati scolastici e all'ansia e alla depressione osservate da Kagan.

Vi sono indicazioni del fatto che i bambini estremamente timorosi possono essere predisposti alle malattie più dei loro coetanei. Quando si sentono gravemente inibiti in situazioni poco familiari, molti di essi manifestano una sovrapproduzione cronica di ormoni dello stress, fra cui il cortisolo, che è sintetizzato dalle ghiandole surrenali. In una situazione di pericolo, questi ormoni sono fondamentali perché garantiscono ai muscoli l'energia necessaria per «combattere o fuggire», ma alcune osservazioni indicano che un innalzamento a lungo termine del livello di ormoni dello stress può contribuire all'insorgere di ulcere gastriche e disturbi cardiovascolari.

Inoltre, per meccanismi sconosciuti, i bambini timorosi e i loro familiari hanno più probabilità di altri di soffrire di disturbi allergici. Infine, nei roditori e nei primati non umani, livelli costantemente elevati di cortisolo sono stati associati a un aumento della vulnerabilità dei neuroni dell'ippocampo o ai danni provocati da parte di altre sostanze (è noto che questa regione cerebrale è correlata alla memoria, alla motivazione e alle emozioni). È probabile che anche nell'uomo i neuroni vengano influenzati in modo simile, benché non se ne abbiano prove dirette.

Quando iniziammo i nostri studi, approssimativamente 10 anni fa, Shelton e io sapevamo che avremmo dovuto per prima cosa trovare quali siano gli stimoli che suscitano paura e identificare quali comportamenti riflettano diversi tipi di ansia. Disponendo di queste informazioni, avremmo allora potuto stabilire a quale età le scimmie comincino ad associare specificamente particolari comportamenti difensivi a determinati stimoli. Stabilendo anche quali regioni cerebrali giungano a maturità nello stesso arco di tempo, avremmo potuto altresì identificare le aree che presiedono alla regolazione della paura e dei comportamenti a essa correlati.

I nostri esperimenti furono eseguiti presso il Wisconsin Regional Primate Research Center e lo Harlow Primate Laboratory, entrambi all'Università del Wisconsin. Abbiamo distinto diversi comportamenti esponendo scimmie di 6-12 mesi di età a tre situazioni correlate. Nella condizione di solitudine, l'animale era separato dalla madre e lasciato da solo in una gabbia per 10 minuti. Nella condizione di mancato contatto oculare, una persona stava immobile all'esterno della gabbia evitando di guardare il piccolo, rinchiuso da solo. Nella condizione di osservazione continua, la persona stava ancora immobile vicino alla gabbia, ma, assumendo un'espressione neutra, fissava direttamente l'animale. Queste situazioni non incutono di per sé maggiore paura di quella che le scimmie possono frequentemente incontrare in natura, o che i bambini sperimentano ogni volta che vengono lasciati alla scuola materna.

Nella condizione di solitudine, la maggior parte delle scimmie diveniva molto attiva e frequentemente «tubava»: emetteva cioè suoni piuttosto melodiosi, che vengono prodotti arricciando le labbra e che iniziano su una tonalità bassa per poi divenire più acuti prima di cessare. Più di 30 anni fa, Harry F. Harlow, allora all'Università del Wisconsin, stabilì che, quando un piccolo di scimmia è separato dalla madre, il suo obiettivo principale è il riavvicinamento: esso desidera ritrovare il senso di vicinanza e di sicurezza fornito dalla presenza della madre. I movimenti e i richiami servono ad attrarne l'attenzione.

Invece nella condizione di mancato contatto oculare, che genera maggiore paura, le scimmie riducevano notevolmente la loro attività e qualche volta si irrigidivano rimanendo completamente immobili per periodi prolungati. Quando un piccolo nota un possibile predatore, il suo obiettivo non è più quello di attrarre la madre, ma quello di rendersi poco evidente. La cessazione dei movimenti e l'irrigidimento - risposte comuni in molte specie - riducono la probabilità di attacco.

Se il piccolo ritiene di essere stato individuato, il suo obiettivo diventa invece quello di scongiurare l'attacco; perciò la condizione di osservazione continua induceva una terza serie di risposte. Le scimmie assumevano atteggiamenti ostili di tipo diverso, tra i quali «abbaiare» (emettere forzatamente aria attraverso le pieghe vocali per produrre una sorta di ringhio), ricambiare lo sguardo, assumere espressioni di minaccia (*si veda l'illustrazione in questa pagina*), scoprire i denti e scuotere la gabbia.

Qualche volta gli animali alternavano gli atteggiamenti minacciosi con mani-

Una femmina di reso si mette in allarme (*a destra*) quando un'altra scimmia (*a sinistra*) si avvicina al suo piccolo. La paura della madre è resa evidente dall'espressione di «minaccia»: la bocca spalancata e lo sguardo penetrante servono a intimidire l'eventuale assalitore.

Tre condizioni sperimentali suscitano distinti comportamenti associati alla paura in piccoli di reso che abbiano superato i due mesi di età. Isolati in una gabbia (*a sinistra*), i piccoli diventano molto agitati e «tubano» per richiamare la madre. Se una persona si avvicina, ma evita il contatto oculare (*al centro*), le scimmie tentano di non farsi notare, o rimanendo immobili (irrigidendosi) o nascondendosi dietro il contenitore del cibo. Se l'intruso guarda fissamente gli animali (*a destra*), essi assumono un atteggiamento aggressivo.

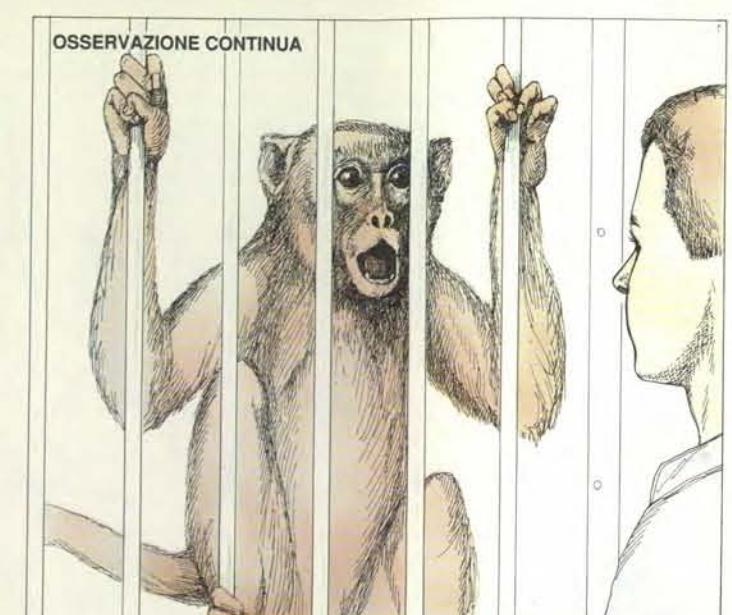
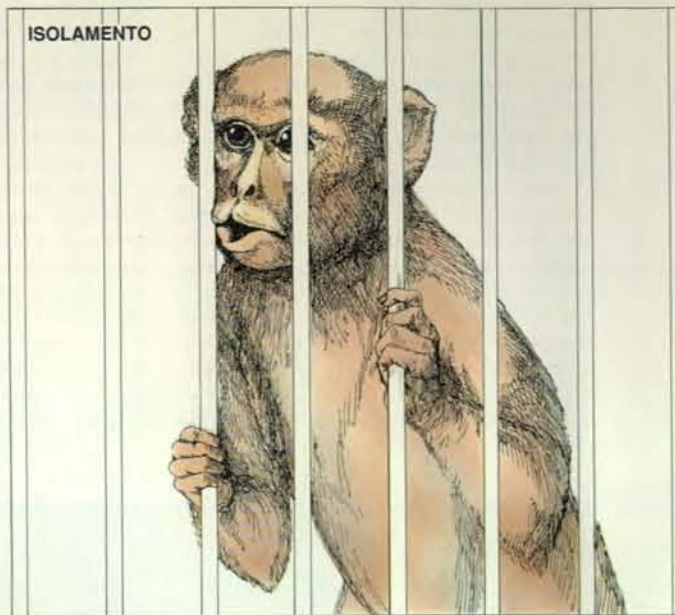
festazioni di sottomissione, quali smorfie di paura, che appaiono quasi come cauti sorrisi, o il digrignare i denti. In questa condizione, inoltre, gli animali tubavano più di quanto facessero in solitudine. (Come vedremo, abbiamo recentemente dedotto che questo tipo di richiamo, nella condizione di osservazione continua, abbia una funzione leggermente diversa da quella della condizione di solitudine.)

Le scimmie, tra l'altro, non sono gli unici animali a essere allarmati dallo sguardo fisso e a usarlo a loro volta per intimidire i predatori; anche granchi, lucertole, uccelli percepiscono l'osservazione continua come una minaccia. Alcuni pesci e insetti hanno evoluto a scopo protettivo chiazze che assomigliano a occhi e che hanno lo scopo di scongiurare gli attacchi o almeno di dirigerli verso parti non vitali del corpo. In India, molti contadini portano maschere che rappresentano un volto sulla parte posteriore del capo per scoraggiare le tigri dall'aggredirli alle spalle. Si è dimostrato che anche gli esseri umani sono sensibili a uno sguardo diretto: l'attività cerebrale di un individuo osservato fissamente aumenta, e le persone ansiose o depresse tendono a evitare il contatto oculare diretto.

Avendo identificato tre insiemi di comportamenti difensivi, ci accingemmo a stabilire quando i piccoli di scimmia cominciano a servirsene efficacemente. Diversi filoni di ricerca ci hanno condotto a ipotizzare che la capacità di fare queste scelte cominci a manifestarsi quando il piccolo ha circa due mesi di età. Per esempio, nei reso, le madri generalmente permettono ai figli di allontanarsi assieme ai loro coetanei intorno a quell'età, presumibilmente perché ritengono che i piccoli siano già in grado di proteggersi abbastanza bene. Sapevamo anche che intorno alle 10 settimane di età i piccoli rispondono con emozioni diverse a specifiche espressioni del viso di altre scimmie, un'indicazione del fatto che almeno alcuni dei circuiti innati o delle capacità apprese finalizzati alla distinzione dei segnali minacciosi sono già acquisiti.

Per stabilire quale sia il periodo critico dello sviluppo, abbiamo esaminato quattro gruppi di scimmie di età compresa fra pochi giorni e 12 settimane. Abbiamo separato i piccoli dalle madri e abbiamo lasciato che si abituassero a una gabbia non familiare; soltanto successivamente li abbiamo sottoposti alle condizioni di solitudine, di mancato contatto oculare e di osservazione continua, registrando i risultati su una videocassetta.

Operando in questo modo abbiamo scoperto che le scimmie appartenenti al gruppo di età inferiore (da pochi giorni a due settimane) manifestavano comportamenti difensivi, ma mancavano di coordinazione dei movimenti e sembravano agire in maniera casuale, come se non fossero state consapevoli della presenza o dello sguardo dell'intruso umano. I piccoli dei due gruppi di età intermedi parevano avere un buon controllo motorio, ma le loro azioni non sembravano correlate alla condizione dell'esperimento. Questa osservazione indica che la coordinazione motoria non è un fattore determinante per la selettività nella risposta.

Soltanto gli animali del gruppo di età superiore (da 9 a 12 settimane) si comportavano in maniera differente nelle diverse situazioni, e le loro reazioni erano appropriate e sostanzialmente identiche a quelle delle scimmie adulte. Il periodo compreso fra 9 e 12 settimane di età può quindi essere considerato di importanza fondamentale in merito alla comparsa delle capacità di adattare le proprie attività difensive a situazioni mutevoli.

Studi condotti da altri ricercatori soprattutto sui roditori hanno indicato che le manifestazioni della paura sono regolate da tre aree cerebrali interconnesse. La nostra deduzione è che queste regioni diventino funzionalmente mature durante il periodo compreso fra le 9 e le 12 settimane e diano quindi origine alla reattività selettiva che abbiamo osservato. Una di queste regioni è la corteccia prefrontale, che occupa gran parte delle aree esterne e laterali della corteccia cerebrale nel lobo frontale (*si veda l'illustrazione a pagina 70*). Si ritiene che la corteccia prefrontale, che è un'area deputata a funzioni cognitive ed emotive, partecipi all'interpretazione degli stimoli sensoriali e sia probabilmente il sito dove viene valutato un pericolo potenziale.

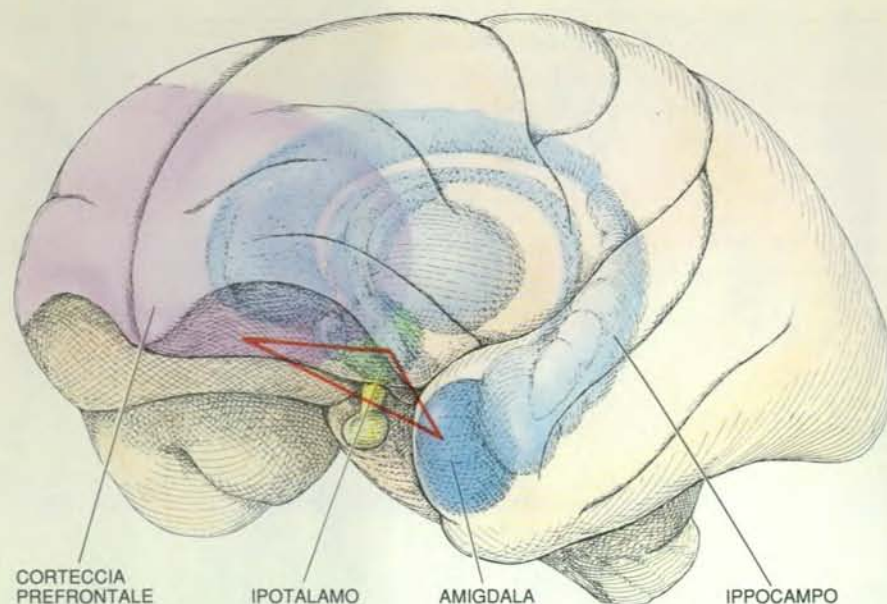
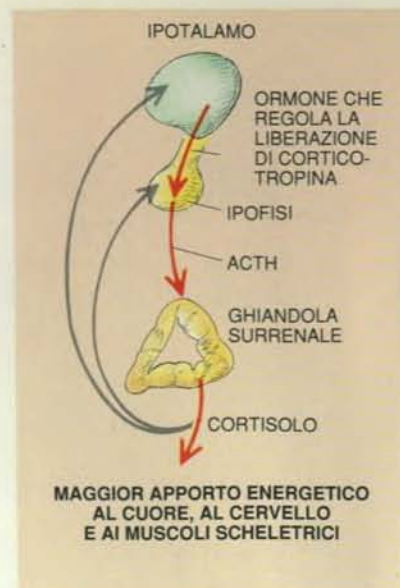
La seconda regione è l'amigdala, che appartiene all'area cerebrale primitiva chiamata sistema limbico (comprendente l'ippocampo). Si è proposto che il sistema limbico in generale e l'amigdala in particolare siano coinvolti nel generare la paura.

La terza regione è l'ipotalamo; situato alla base del cervello, è un costituente del cosiddetto sistema ipotalamico-ipofisario-surrenale. In risposta a segnali di stress provenienti da altre regioni cerebrali, quali il sistema limbico e altre aree corticali, l'ipotalamo secerne l'ormone che promuove la liberazione di corticotropina. Questa piccola proteina induce l'ipofisi, una ghiandola situata sulla faccia inferiore dell'ipotalamo, a secernere ormone adrenocorticotropo (ACTH), il quale, a sua volta, stimola le ghiandole surrenali a liberare cortisolo (o idrocortisone), uno steroide adrenocorticale che prepara l'organismo a difendersi.

I dati neuroanatomici raccolti in altri laboratori hanno dato sostegno all'ipotesi da noi formulata, ossia che la maturazione di queste regioni cerebrali consenta reazioni selettive durante il periodo compreso tra 9 e 12 settimane di età. Per esempio si è dimostrato che in questo intervallo di tempo la formazione di sinapsi (punti di contatto fra neuroni) raggiunge la massima intensità nella corteccia prefrontale e nel sistema limbico (compresa l'amigdala), oltre che nella corteccia motoria e visiva e in altre aree sensoriali. Patricia S. Goldman-Rakic della Yale University ha anche stabilito che nelle scimmie reso la maturazione della corteccia prefrontale si accompagna al manifestarsi della capacità di regolare il proprio comportamento in base all'esperienza. Questa conquista è necessaria per affrontare con successo il pericolo.

La maturazione della corteccia prefrontale sembra di peculiare importanza anche nell'uomo per consentire di distinguere i segnali di minaccia. Harry T. Chugani e collaboratori dell'Università della California a Los Angeles hanno dimostrato che l'attività della corteccia prefrontale aumenta nei bambini fra i 7 e i 12 mesi di età. In questo intervallo - che sembra analogo al periodo durante il quale le scimmie cominciano a reagire selettivamente alla paura - i bambini cominciano a mostrare una evidente paura degli estranei. Cominciano anche ad acquisire la capacità di regolare il proprio livello di paura basandosi sulla interpretazione dell'espressione del viso di un genitore.

Ma che dire dell'ipotalamo, la terza

I comportamenti indotti in laboratorio dalle condizioni di isolamento, di mancato contatto oculare e di osservazione continua - «tubare», irrigidirsi e mostrare minacciosamente i denti - si osservano anche in natura nei piccoli e negli adulti spaventati. Le fotografie sono state eseguite a Cayo Santiago, un'isola al largo di Puerto Rico.

Tre regioni cerebrali interconnesse da vie nervose (*indicate dal triangolo in rosso*) sono fondamentali per regolare i comportamenti correlati alla paura. La corteccia prefrontale fornisce la valutazione del pericolo, mentre l'amigdala è una componente importante del sistema limbico che governa le emozioni (*in blu chiaro*); l'ipotalamo, in risposta a segnali provenienti dalla corteccia prefrontale, dall'amigdala e dall'ippocampo, regola la secrezione di ormoni (*frecce rosse nel riquadro*) che attivano le risposte motorie a situazioni di minaccia. (Le frecce grige evidenziano il ruolo inibitorio del cortisolo.)

regione cerebrale che a nostro parere potrebbe contribuire alla regolazione del comportamento correlato alla paura? Le ricerche pubblicate a tutt'oggi non offrono informazioni significative sul suo sviluppo né sullo sviluppo dell'intero sistema ipotalamico-ipofisario-surrenale nelle scimmie. Le nostre indagini, tuttavia, hanno indicato che l'intero sistema matura in parallelo con la corteccia prefrontale e il sistema limbico.

Per questi esperimenti abbiamo utilizzato l'ormone ipofisario ACTH come marcatore della funzionalità del sistema. Abbiamo esaminato di nuovo quattro gruppi di piccoli di età compresa fra pochi giorni e 12 settimane. Per ciascun soggetto, abbiamo misurato i livelli ematici di ACTH quando il piccolo era ancora con la madre, in modo da ottenere un valore di base. Abbiamo inoltre misurato i livelli di ACTH in campioni di sangue prelevati 20 minuti dopo che il piccolo era stato separato dalla madre. I livelli ormonali sono aumentati in tutti e quattro i gruppi di età durante la separazione, ma il dislivello è stato marcato solo nelle scimmie di età più elevata (da 9 a 12 settimane).

La risposta relativamente scarsa degli animali più giovani, in particolare di quelli che avevano meno di due settimane di età, concorda con osservazioni eseguite su piccoli di ratto, nei quali la risposta agli ormoni dello stress è analogamente modesta durante le prime due settimane di vita. Lo sviluppo del sistema di ormoni dello stress nei roditori e nei primati potrebbe essere ritardato nelle primissime fasi di vita per proteggere i neuroni dagli effetti potenzialmente dannosi del cortisolo.

Avendo confermato che il sistema ipotalamico-ipofisario-surrenale diventa funzionalmente maturo fra 9 e 12 settimane, continuammo le nostre ricerche per determinare se i livelli di cortisolo e di ACTH possono in parte giustificare le differenze individuali nel comportamento difensivo. Eravamo anche curiosi di sapere se le risposte dei piccoli erano simili a quelle delle loro madri; una correlazione tra le risposte, infatti, indicherebbe la possibilità di stabilire il contributo relativo dei caratteri ereditati o appresi negli atteggiamenti di paura. Abbiamo esaminato soprattutto la propensione a irrigidirsi, che, come avevamo riscontrato in precedenza, era un tratto stabile nei nostri soggetti.

In una serie di studi, abbiamo misurato i livelli basali di cortisolo in scimmie di età compresa fra quattro mesi e un anno, e successivamente abbiamo osservato per quanto tempo i piccoli si irrigidivano nella situazione di mancato contatto oculare. Le scimmie che inizialmente avevano livelli relativamente bassi di cortisolo si irrigidivano per tempi più brevi di quelle con livelli di cortisolo elevati, una caratteristica che abbiamo notato anche in studi separati di femmine adulte. In altri esperimenti abbiamo osservato che, nel corso del primo anno di età, gli individui giovani diventano sempre più simili alle loro madri sotto il profilo sia ormonale sia comportamentale. A circa cinque mesi l'innalzamento del livello di ACTH indotto dallo stress è parallelo a quello delle madri, e a un anno di età la durata dell'irrigidimento nella condizione di mancato contatto oculare corrisponde anch'essa a quella della madre.

È significativo che alcuni di questi risultati riflettano osservazioni compiute sull'uomo. Spesso i bambini estremamente inibiti hanno genitori che soffrono di ansia. Inoltre Kagan e colleghi hanno trovato che i livelli basali di cortisolo sono indicativi delle reazioni di questi bambini a una situazione che incute paura. Essi hanno misurato la concentrazione di cortisolo nella saliva dei bambini quando questi erano a casa (e probabilmente erano più rilassati) e poi li hanno messi di fronte a una situazione insolita in laboratorio: elevati livelli basali di cortisolo erano associati a una maggiore inibizione nella situazione non familiare.

Queste somiglianze fra uomo e scimmie lasciano supporre che queste ultime possano essere considerate modelli attendibili della risposta emotiva umana. Il collegamento fra livelli basali di cortisolo e durata dell'irrigidimento o dell'inibizione fa pensare anche che i livelli degli ormoni dello stress determinino in che misura sia appropriata la reazione sia degli animali sia degli esseri umani di fronte alla paura. (Questo effetto può essere parzialmente mediato dall'ippocampo, dove la concentrazione di recettori del cortisolo è elevata.) Inoltre, la somiglianza delle risposte ormonali e comportamentali in madri e figli implica che l'ereditarietà genetica possa predisporre alcuni individui a una estrema timorosità, sebbene non sia possibile escludere un significativo contributo dell'esperienza.

Un piccolo (*a sinistra*) si è allontanato un poco dalla madre (*al centro*) e prova ad assumere un'espressione minacciosa nella speranza di tenere lontano l'autore che lo sta fotografando. I reso acquisiscono la capacità di adattare il proprio comportamento al tipo e alla gravità della minaccia fra le 9 e le 12 settimane di età, probabilmente perché è in questo periodo che determinate vie nervose della corteccia prefrontale, dell'amigdala e dell'ipotalamo raggiungono la maturità funzionale.

Non si può ancora dire fino a che punto l'attività del sistema ipotalamico-ipofisario-surrenale controlli e sia controllata da altre regioni cerebrali che regolano la scelta dei comportamenti difensivi. Abbiamo tuttavia cominciato a identificare circuiti, o sistemi, neurochimici cerebrali distinti che influenzano comportamenti diversi. I due sistemi che abbiamo studiato più a fondo sembravano in un primo tempo avere funzioni del tutto separate. Ma osservazioni più recenti paiono dimostrare che i controlli sul comportamento difensivo siano alquanto più complicati di quanto sembrasse all'inizio.

Abbiamo raccolto i nostri primi dati tre anni fa, trattando scimmie di 6-12 mesi con due diverse classi di sostanze neuroattive: oppiacei (sostanze contenenti alcaloidi dell'oppio o loro derivati come la morfina) e benzodiazepine (psicofarmaci ad azione ansiolitica e sedativa come il diazepam). Abbiamo scelto di impiegare gli oppiacei e le benzodiazepine perché i neuroni che liberano o assumono queste sostanze sono abbondanti nella corteccia prefrontale, nell'amigdala e nell'ipotalamo. Si sa che gli oppiacei hanno controparti naturali, o endogene, chiamate endorfine ed encefaline, che fungono da neurotrasmettitori; quando queste sostanze endogene sono liberate da certi neuroni, si legano a molecole recettrici su altre cellule nervose e ne aumentano a dismisura l'attività. Sono stati identificati i recettori delle benzodiazepine, ma non si è ancora riusciti a isolare molecole endogene simili a questi composti.

Ancora una volta i nostri macachi furono sottoposti alle condizioni di solitudine, di mancato contatto oculare e di osservazione continua. Somministrammo i farmaci prima che i piccoli fossero separati dalle madri e poi procedemmo alla registrazione del comportamento degli animali. La somministrazione di morfina ridusse la reazione del «tubare» mostrata normalmente nelle condizioni di solitudine e di osservazione continua. Questo comportamento venne invece accentuato dal naloxone, un composto che si lega ai recettori degli oppiacei, ma blocca l'attività della morfina e degli oppioidi endogeni. Tuttavia la morfina e il naloxone non ebbero alcun effetto sulla frequenza del ringhio indotto dall'osservazione continua e su altri comportamenti ostili, né influenzarono la durata dell'irrigidimento nella situazione di mancato contatto oculare. Ne abbiamo concluso che le vie nervose che impiegano oppioidi regolano in primo luogo i comportamenti di richiamo (come quelli indotti dal disagio per la separazione dalla madre), ma queste vie sembrano avere scarso effetto sulle risposte a minacce dirette.

La benzodiazepina da noi studiata - il diazepam - ha prodotto risultati opposti. Non ha avuto alcun effetto sul «tubare», ma ha ridotto notevolmente l'irrigidimento, il ringhio e altri atteggiamenti ostili. Per questa ragione le vie nervose che impiegano benzodiazepine sembrano influenzare soprattutto le risposte a minacce dirette, ma non i comportamenti di richiamo.

	TUBARE	IRRIGIDIRSI	ABBAIARE
MORFINA (OPPIOIDE)	DIMINUISCE	NESSUN EFFETTO	NESSUN EFFETTO
NALOXONE (ANTAGONISTA DEGLI OPPIOIDI)	AUMENTA	NESSUN EFFETTO	NESSUN EFFETTO
DIAZEPAM (BENZODIAZEPINA)	NESSUN EFFETTO	DIMINUISCE	DIMINUISCE

Alcuni anni fa si sono valutati gli effetti sul «tubare», sull'irrigidirsi e sull'«abbaiare» di tre sostanze che agiscono sui neuroni sensibili agli oppioidi (*le due file in alto*) o alle benzodiazepine (*la fila in basso*). I risultati fanno ritenere che le vie cerebrali sensibili agli oppioidi controllino i comportamenti che tendono a richiamare la madre (come spesso accade con il «tubare»), mentre le vie sensibili alle benzodiazepine regolino le risposte (come l'irrigidirsi e l'«abbaiare») a situazioni minacciose. Nuovi dati confermano questa conclusione, pur complicando il quadro.

Siamo ancora convinti che le vie degli oppioidi e delle benzodiazepine abbiano, a livello basilare, queste funzioni separate. Tuttavia il semplice modello che avevamo messo a punto inizialmente è diventato più interessante quando abbiamo preso in considerazione altri due farmaci: una benzodiazepina, l'alprazolam, e la beta-carbolina, una sostanza che si lega ai recettori delle benzodiazepine, ma provoca un aumento dell'ansia e tipicamente produce effetti opposti a quelli del diazepam e dei suoi analoghi.

Quando abbiamo somministrato alprazolam in dosi che diminuiscono l'ansia in misura sufficiente da ridurre l'irrigidimento, questa sostanza, come il diazepam, ha reso minima l'ostilità nella situazione di osservazione continua; la beta-carbolina ha invece incrementa-

Una madre tranquilla (*a sinistra*) non reagisce quasi alla presenza dell'autore con la sua macchina fotografica, mentre una più sensibile si spaventa (*a destra*), come dimostra la sua «smorfia di paura». L'autore spera che lo studio delle basi nervose di questi differenti comportamenti nelle scimmie faciliti lo sviluppo di nuove terapie contro l'ansia nell'uomo.

to l'ostilità. Nessuna sorpresa, dunque. Tuttavia, al contrario del diazepam, questi farmaci hanno influito anche sul «tubare», che consideravamo un comportamento di richiamo (controllato dagli oppioidi) e non correlato alle minacce (controllate dalle benzodiazepine); l'effetto di entrambi i composti è stato di ridurre questo comportamento. Non possiamo spiegare la somiglianza degli effetti, ma abbiamo formulato qualche ipotesi per spiegare come mai sostanze che agiscono sui recettori delle benzodiazepine possano avere influenza su questo comportamento.

Può darsi che, contrariamente alle nostre prime supposizioni, le vie delle benzodiazepine possano in realtà controllare anche i comportamenti di richiamo; tuttavia riteniamo più plausibile una seconda interpretazione. Nella condizione di osservazione continua, il «tubare» può non solo riflettere un bisogno di riavvicinamento (il desiderio del conforto dato dalla madre), ma, a volte, può essere anche una pressante invocazione d'aiuto immediato indotta da una minaccia. Uno stesso comportamento, quindi, può avere due funzioni differenti ed essere controllato da vie neurochimiche diverse. (Questa conclusione ha recentemente trovato una conferma quando ho cercato di fotografare nel suo ambiente naturale un piccolo reso che era rimasto separato dalla madre. Il suo «tubare» intenso e persistente ha attratto la madre insieme con un gruppo di difensori. La strategia ha funzionato: ho dovuto infatti battere in ritirata.)

Più in generale, i nostri studi chimici ci hanno condotto a supporre che in situazione di stress entrino in funzione sia i circuiti sensibili agli oppioidi sia quelli sensibili alle benzodiazepine; il relativo grado di attività cambia con le caratteristiche della situazione stressante. Al variare del contributo di ciascuna via, si modificano anche i comportamenti manifestati.

Non è ancora chiaro come funzionino esattamente i neuroni delle vie degli oppioidi e delle benzodiazepine e in che modo possano cooperare. Ma uno scenario plausibile è il seguente: quando una giovane scimmia è separata dalla madre, i neuroni che liberano oppioidi, e di conseguenza sono sensibili a queste sostanze, si inattivano. Questo fenomeno suscita un intenso desiderio della madre e una sensazione di vulnerabilità. L'abbassamento di attività delle vie nervose sensibili agli oppioidi permette ai sistemi motori cerebrali di produrre il «tubare». Quando appare un potenziale predatore, i neuroni che secernono benzodiazepine endogene vengono in una certa misura inibiti. Questo cambiamento, a sua volta, produce una notevole ansia e la comparsa dei comportamenti e delle risposte ormonali che accompagnano la paura. Al crescere della sensazione di allarme, le aree motorie si preparano alla lotta o alla fuga. Il sistema delle benzodiazepine può anche influenzare il sistema degli oppioidi, alterando quindi la reazione del «tubare» in situazioni minacciose.

Attualmente stiamo perfezionando il nostro modello delle funzioni cerebrali sperimentando altri composti che si legano ai recettori degli oppioidi e delle benzodiazepine. Stiamo anche esaminando le risposte comportamentali a sostanze, come la serotonina, che agiscono su altri recettori. (I recettori della serotonina si trovano in molte regioni cerebrali coinvolte nella manifestazione della paura.) Stiamo infine studiando l'attività di sostanze che controllano direttamente la produzione di ormoni dello stress, fra i quali, in particolare, l'ormone che stimola la liberazione di corticotropina, presente in tutto il cervello e non solo nell'ipotalamo.

In collaborazione con Richard J. Davidson, dell'Università del Wisconsin, Shelton e io abbiamo di recente identificato almeno una regione cerebrale dove il sistema delle benzodiazepine esercita i propri effetti. Davidson ha dimostrato che la corteccia prefrontale dell'emisfero destro è insolitamente attiva nei bambini estremamente inibiti. Ci siamo quindi chiesti se avremmo osservato la stessa asimmetria nelle scimmie timorose, e se i farmaci che riducono i comportamenti correlati alla paura negli animali avrebbero attenuato l'attività della regione destra del cervello.

In questo caso abbiamo utilizzato come fonte di stress una moderata restrizione del movimento. Come prevedevamo, si è avuto un maggiore incremento dell'attività dei neuroni nella corteccia frontale destra che non nella sinistra. Inoltre, quando abbiamo somministrato diazepam in dosi tali da diminuire l'ostilità, l'attività elettrica indotta dalla situazione di stress è tornata al livello normale. In altri termini, il sistema delle benzodiazepine influisce, almeno in parte, sul comportamento difensivo agendo sulla corteccia prefrontale destra.

Queste scoperte hanno implicazioni terapeutiche. Se il cervello dell'uomo e delle scimmie funziona in maniera analoga, i nostri dati starebbero a indicare che le benzodiazepine possono dimostrarsi molto utili negli adulti e nei bambini che presentano elevata attività elettrica della corteccia prefrontale destra. A

P.A.M. Dirac e la bellezza della fisica

Preferiva una teoria bella a una poco elegante, ma di alto contenuto empirico perché, come soleva asserire, i fatti cambiano; e dimostrò che il suo punto di vista era corretto prevedendo l'esistenza dell'antimateria

di R. Corby Hovis e Helge Kragh

I fisici celebri in visita all'Università di Mosca sono invitati a lasciare una dichiarazione per i posteri su una lavagna. Niels Bohr, padre della teoria atomica quantistica, vi iscrisse il motto del famoso principio di complementarità: *Contraria non contradictoria sed complementa sunt* (gli opposti non sono contraddittori, ma complementari). Hideki Yukawa, pioniere della moderna teoria dell'interazione nucleare forte, tracciò la frase: «L'essenza della natura è semplice». Paul Adrien Maurice Dirac scelse l'epigrafe: «Una legge fisica deve possedere bellezza matematica».

Esattamente trent'anni fa Dirac scrisse, per un numero di «Scientific \merican»: «Dio è un matematico di altissimo livello e si è servito di un formalismo estremamente evoluto per costruire l'universo». Ispirato dal pensiero di Albert Einstein ed Hermann Weyl, Dirac, più di ogni altro fisico moderno, interpretò il concetto di «bellezza matematica» come una caratteristica intrinseca della natura e lo assunse a guida metodologica dell'indagine scientifica fino ad asserire che «una teoria dotata di bellezza matematica ha più probabilità di essere corretta che non una sgradevole che concordi con qualche dato sperimentale».

L'attenzione di Dirac per l'estetica e la logica della fisica matematica, associata a una reticenza e a un'introversione leggendarie, lo rese una delle figure più enigmatiche tra i grandi scienziati di questo secolo. Purtroppo sembra che il suo estremo razionalismo lo abbia spesso condotto in sterili vicoli ciechi dopo gli straordinari successi conseguiti nei primi anni di attività. Tra i 23 e i 31 anni di età, infatti, Dirac formulò un originale e potente modello della meccanica quantistica, una teoria quantistica sull'emissione e sull'assorbimento di radiazione elettromagnetica da parte dei sistemi atomici (una versione primitiva ma importante dell'elettrodinamica quantistica), l'equazione d'onda relativistica dell'elettrone, il concetto di antimateria e una teoria dei monopoli magnetici. Invece pochi dei suoi successivi contributi furono di valore assoluto, e nessuno ebbe il carattere rivoluzionario dei primi lavori.

Dirac nacque nel 1902 a Bristol, nel Regno Unito, secondo di tre figli di una famiglia che oggi sarebbe giudicata oppressiva. La «bestia nera» era il capofamiglia, Charles Adrien Ladislas Dirac, che aveva sposato Florence Hannah Holten, figlia di un comandante di vascello, dopo essere emigrato dalla Svizzera nel Regno Unito intorno al 1890. Charles visse lavorando come insegnante di francese presso il Merchant Venturers' Technical College di Bristol, dove adottava criteri di disciplina estremamente autoritari. E gli stessi principi di decoro militaresco riservò ai componenti della famiglia Dirac. Lesinando sulle dimostrazioni d'affetto e identificando l'amore paterno con la disciplina, imprigionò di fatto i suoi figli in una tirannia domestica che li isolò dalla vita sociale e culturale. Non riuscendo a ribellarsi, o non volendolo, Paul si rifugiò nel silenzio e prese le distanze dal padre, ma quegli anni infelici lo segnarono per sempre. Quando Charles Dirac morì, nel 1936, egli non ne fu afflitto, al punto da scrivere alla futura moglie: «Adesso mi sento più libero».

Fortunatamente Paul era animato da una ricchezza interiore in cui poteva trovare rifugio, e inoltre mostrava una precoce inclinazione naturale per la matematica. All'età di 12 anni si iscrisse al Merchant Venturers' Technical College. Questa scuola, a differenza di molti altri istituti dell'epoca, non offriva un'educazione classica basata sull'insegnamento del latino e del greco ma presentava moderni programmi di scienze, lingue e materie pratiche. Questi studi furono soddisfacenti per il giovane Dirac perché, come ebbe modo di dire, non «apprezzava il valore della cultura antica». Dopo aver portato a termine la scuola superiore entrò in un altro istituto ubicato nello stesso edificio, l'Engineering College dell'Università di Bristol, ove intraprese studi di ingegneria elettrica non tanto per entusiasmo nei confronti di questa materia, ma perché convinto di meritare così l'approvazione del padre.

Il programma di ingegneria lasciava poco spazio ad argomenti diversi dalla fisica e dalla matematica applicate. A dispetto di queste lacune, Dirac fu affascinato dalle nuove teorie di Einstein sullo spazio, sul tempo e sulla gravità (le teorie della relatività speciale e generale) e ben presto le conobbe a fondo.

Quando Dirac si laureò con il massimo dei voti, nel 1921, il mondo del lavoro era pesantemente condizionato dalla depressione economica postbellica, ed egli fu salvato da una borsa di studio in matematica a Bristol, dopo la quale proseguì, nell'autunno 1923, con studi universitari di matematica applicata e fisica teorica all'Università di Cambridge. A quel tempo, a Cambridge erano di casa eminenti scienziati - quali Joseph Larmor, J. J. Thomson, Ernest Rutherford, Arthur Stanley Eddington e James Jeans - e astri nascenti del calibro di James Chadwick, Patrick Blackett, Edward A. Milne, Ralph Fowler, Douglas R. Hartree e Peter Kapitza. Affidato alla tutela di Fowler, Dirac apprese la teoria atomica e la meccanica statistica, argomenti che non aveva studiato in precedenza. Di quegli anni rammentò in seguito: «Mi rinchiusi completamente nel lavoro scientifico, e continuai a dedicarmi a esso giorno dopo giorno, salvo le domeniche, in cui riposavo e, se c'era bel tempo, facevo lunghe passeggiate solitarie in campagna».

«Era alto, magro, goffo ed estremamente taciturno» scrisse il fisico e biologo tedesco Walter Elsasser. «Era riuscito a dedicare tutte le proprie energie a un unico interesse preminente. Era quindi un uomo di grandezza straordinaria in un campo, ma con scarso interesse e limitate competenze nelle altre attività umane... In altre parole, era il prototipo della mente matematica superiore; ma, mentre in altri questa qualità coesiste con una grande varietà di interessi, nel caso di Dirac tutto fu volto al compimento della sua storica missione: l'istituzione di una nuova scienza, la meccanica quantistica, alla quale probabilmente contribuì più di chiunque altro.»

Memorabilia

Nel 1931, quando era lettore a Cambridge, Nevill Mott scrisse ai genitori: «Dirac è pressappoco come l'idea che ci si fa di Gandhi. È stato a cena da noi... Era proprio un'ottima cenetta, ma sono sicuro che, se gli avessimo offerto soltanto del *porridge*, non vi avrebbe badato. Va a Copenaghen sulla rotta del Mare del Nord perché pensa di doversi curare dal mal di mare. Non riesce assolutamente a fingere di pensare qualcosa che non pensi veramente. All'epoca di Galileo sarebbe stato un martire pienamente soddisfatto».

Eugene Wigner

J. Robert Oppenheimer

Una volta Dirac partecipò a un pranzo con Eugene Wigner e Michael Polanyi. Si scatenò una vivace discussione sulla scienza e la società, durante la quale Dirac non disse una parola. Richiesto della sua opinione, rispose: «Ci sono sempre più persone disposte a parlare che persone disposte ad ascoltare».

Un fisico francese, che parlava a stento l'inglese, fece visita a Dirac. Questi ascoltò pazientemente mentre il collega tentava di trovare i vocaboli inglesi corretti per esporre le proprie opinioni. A un tratto la sorella di Dirac entrò nella stanza chiedendo, in francese, qualcosa al fratello, il quale a sua volta rispose in ottimo francese. Naturalmente l'interlocutore si indignò, e proruppe: «Perché non mi avete detto che parlate francese?» E Dirac replicò candidamente: «Perché non me lo avete chiesto».

Nel 1934, passando da Berkeley diretto in Giappone, Dirac incontrò J. Robert Oppenheimer, che gli offrì due libri da leggere durante il viaggio. Dirac, educatamente, rifiutò, dicendo che la lettura interferisce col pensiero. Una volta il fisico russo Peter Kapitza diede a Dirac una traduzione inglese dell'opera di Fëdor Dostojevskij *Delitto e castigo*. Dopo qualche tempo Kapitza chiese a Dirac se avesse apprezzato il libro. Il suo solo commento fu: «È molto gradevole, ma in un capitolo l'autore ha commesso un errore, descrivendo due volte nello stesso giorno il sorgere del Sole». Dietro consiglio, Dirac lesse anche *Guerra e pace* di Lev Nikolajevič Tolstoj: gli occorsero due anni.

Peter Kapitza

Dirac sfuggiva la pubblicità. Sulle prime era intenzionato a rifiutare il premio Nobel. Il giorno per il quale era fissata la sua nomina alla Cattedra lucasiana, se ne andò allo zoo per evitare le troppe congratulazioni. Rifiutò tutti i titoli onorari; tuttavia molti gli furono assegnati in sua assenza e, evidentemente, senza il suo consenso.

All'inizio degli anni cinquanta, a Cambridge, Dirac fu assegnato come relatore a Dennis Sciama per la tesi di dottorato. Un giorno Sciama entrò entusiasticamente nell'ufficio di Dirac, strepitando: «Professore, ho appena trovato un modo per correlare la formazione delle stelle alle questioni cosmologiche. Posso parlarvene?» Dirac replicò: «No». Fine della conversazione. Dirac non sembrava rendersi conto che la sua concisione e il suo candore potevano essere percepiti come maleducazione e impudenza.

Quando Dirac partecipava a una conferenza, si sforzava di presentare i suoi testi con la massima lucidità e immediatezza. E considerava illogico cambiare le frasi scelte con tanta cura solo perché non erano state ben afferrate. Più di una volta qualcuno, tra il pubblico, gli chiese di ripetere un passaggio che non aveva capito, il che sottintendeva che l'ascoltatore avrebbe gradito una spiegazione più ampia. In questi casi Dirac ripeteva ciò che aveva detto in precedenza, usando esattamente le stesse parole.

Nel 1977 Dirac scrisse: «Tra tutti i fisici che ho conosciuto, penso che Erwin Schrödinger sia uno di quelli che ho sentito più vicini a me. Mi sono trovato d'accordo con Schrödinger più facilmente che con chiunque altro. Credo che la ragione di questo risieda nel fatto che entrambi abbiamo una grandissima considerazione per la bellezza matematica delle leggi fisiche... Era una specie di atto di fede, per noi, che qualsiasi equazione deputata a descrivere le leggi fondamentali della Natura dovesse avere, in sé, una grande bellezza matematica».

Erwin Schrödinger

Sei mesi dopo il suo arrivo all'ateneo pubblicò il primo articolo scientifico e nei due anni seguenti ne produsse un'altra decina. All'epoca in cui completò la dissertazione di dottorato, nel maggio 1926, aveva già scoperto una originale formulazione della meccanica quantistica e tenuto il primo corso di meccanica quantistica che si sia istituito in un'università britannica. Dopo soli 10 anni dal suo ingresso a Cambridge, appena trentunenne, sarebbe stato insignito del premio Nobel per la fisica per la sua «scoperta di nuove fertili forme della teoria atomica... e per le sue applicazioni».

Gli otto anni di straordinaria produzione scientifica della vita di Dirac ebbero inizio nell'agosto 1925, allorché ricevette da Fowler le bozze di un articolo in corso di pubblicazione di cui era autore un giovane fisico teorico tedesco, Werner Heisenberg (si veda l'articolo *Heisenberg, l'indeterminazione e la rivoluzione quantistica* di David C. Cassidy in «Le Scienze» n. 287, luglio 1992). Quell'articolo esponeva i fondamenti matematici di una rivoluzionaria teoria dei fenomeni atomici che sarebbe presto diventata famosa con il nome di meccanica quantistica. Dirac comprese immediatamente che il lavoro di Heisenberg apriva una strada completamente nuova nella ricerca sul mondo dei fenomeni a scala ultramicroscopica. Nel corso dell'anno successivo Dirac riformulò le fondamentali intuizioni di Heisenberg in una originale teoria della meccanica quantistica che prese il nome di algebra dei *q*-numeri, dal termine da lui impiegato per indicare grandezze fisiche «osservabili» come la posizione, la quantità di moto o l'energia.

Sebbene la sua opera avesse meritato a Dirac ampi e rapidi riconoscimenti, molti dei suoi risultati furono contemporaneamente ottenuti in Germania da un folto gruppo di teorici, tra i quali Heisenberg, Max Born, Wolfgang Pauli e Pascual Jordan, con i quali Dirac ingaggiò un'aperta competizione. Born, Heisenberg e Jordan elaborarono lo schema iniziale di Heisenberg in termini di matematica matriciale. In seguito, nella primavera del 1926, il fisico austriaco Erwin Schrödinger propose un'altra teoria quantistica, la meccanica ondulatoria, che conduceva agli stessi risultati delle ben più astratte teorie formulate da Heisenberg e Dirac e si prestava più facilmente al calcolo. Molti fisici supposero che i tre modelli altro non fossero che semplici rappresentazioni particolari di una più generale teoria della meccanica quantistica.

Durante un soggiorno di sei mesi presso l'Istituto di fisica teorica di Copenaghen, Dirac intuì la teoria generale che tanti scienziati avevano auspicato, un'ossatura che includeva tutti gli schemi particolari e forniva precise regole per trasformarli l'uno nell'altro. La «teoria delle trasformazioni» di Dirac, insieme con un analogo modello elaborato allo stesso tempo da Jordan, pose i fondamenti per ogni successivo sviluppo della meccanica quantistica.

Il 26 dicembre 1927 il fisico britannico Charles G. Darwin (nipote del celebre naturalista) scrisse a Bohr: «Mi trovavo a Cambridge pochi giorni fa, e vi ho incontrato Dirac. Ora dispone di un sistema di equazioni completamente nuovo che descrive correttamente lo spin in tutte le situazioni e sembra essere "la cosa". Le sue sono equazioni differenziali del primo ordine, non del secondo!».

L'equazione di Dirac per l'elettrone era davvero «la cosa», poiché a un tempo soddisfaceva le condizioni della teoria della relatività speciale e rendeva conto dello «spin» dell'elettrone, una grandezza misurabile sperimentalmente che poteva assumere i valori +1/2 o −1/2, «su» o «giù». L'equazione originale di Schrödinger aveva fallito in questa impresa perché non era relativistica, e la sua estensione relativistica, l'equazione di Klein-Gordon, non poteva dar ragione dello spin.

L'uso di equazioni differenziali del primo ordine, tanto degno di nota per Darwin, era decisivo per due ragioni. Innanzitutto Dirac aveva voluto conservare la struttura formale dell'equazione di Schrödinger, che conteneva una derivata prima rispetto al tempo. In secondo luogo, doveva assecondare le restrizioni della teoria della relatività, che collocava lo spazio e il tempo sullo stesso piano. La difficile conciliazione di queste condizioni effettuata da Dirac era al tempo stesso bella e funzionale: quando applicò la nuova equazione al caso di un elettrone che si muove in un campo elettromagnetico, ne scaturì spontaneamente il valore esatto dello spin.

Questa deduzione di una proprietà fisica da principi primi impressionò gli scienziati, che ne parlarono come di un «miracolo», di «una meraviglia assoluta», e si apprestarono ad analizzarne le sottigliezze. Alla fine, questa linea di ricerca portò alla nascita dell'analisi spinoriale e allo sviluppo delle equazioni d'onda relativistiche per particelle con spin diverso da 1/2. In un'altra provvida occasione, quando Dirac e altri scienziati applicarono l'equazione all'atomo di idrogeno, riuscirono a riprodurre con estrema precisione le righe spettrali osservate sperimentalmente. Meno di un anno dopo la pubblicazione, l'equazione di Dirac era già diventata ciò che è rimasta fino ai nostri giorni: una pietra angolare della fisica moderna.

Adoratore della logica matematica, Dirac era anche un maestro di intuizione. Questi tratti intellettuali apparentemente contraddittori non si appalesarono mai tanto apertamente come nella sua teoria delle «buche», elaborata tra il 1929 e il 1931. Questo modello illuminava un intero mondo sfuggito all'attenzione dei fisici.

La teoria prese le mosse dall'intuizione di Dirac che la sua equazione non riguardava soltanto i familiari elettroni di energia positiva, ma anche elettroni dotati di energia negativa. Queste particelle avrebbero mostrato proprietà assolutamente particolari. Inoltre le particelle con energia positiva avrebbero dovuto normalmente cadere in questi stati a energia negativa, provocando il collasso del mondo che ci circonda!

Sul finire del 1929 Dirac trovò una scappatoia per sfuggire all'enigma posto dall'esistenza apparentemente inevitabile di elettroni con energia negativa in natura. Egli immaginò che il vuoto costituisca un «mare» uniforme di stati a energia negativa riempiti da elettroni. Poiché il principio di esclusione di Pauli vieta che due elettroni occupino lo stesso stato quantico, gli elettroni con energia positiva si manterrebbero al di sopra del mare invisibile, a formare gli stati «eccitati» che si osservano in natura. Uno stato eccitato potrebbe anche essere creato fornendo a un elettrone energia positiva a sufficienza per portarlo sopra il mare; nel processo si creerebbe una «buca» in cui potrebbe cadere un altro elettrone con energia negativa. «Queste buche saranno oggetti dotati di energia positiva e perciò dovranno essere trattati come particelle ordinarie» scrisse Dirac all'inizio del 1930.

Ma con quale particella si potrebbe identificare una buca? All'epoca si proponevano due possibili candidati, e Dirac li prese in considerazione entrambi: il protone e l'«elettrone positivo». La prima scelta, il protone, fece sorgere quasi immediatamente due gravi difficoltà. In primo luogo ci si aspetterebbe che un elettrone, occasionalmente, possa compiere un salto energetico verso il basso e andare a colmare una buca, nel qual caso le due particelle si annichilerebbero, con emissione di radiazione gamma; ma queste annichilazioni protone-elettrone non erano mai state osservate. Inoltre, era evidente che il candidato adatto doveva essere identico all'elettrone in ogni caratteristica, fatta eccezione per la carica elettrica, mentre era ben noto che il protone ha massa pari a circa 2000 volte quella dell'elettrone.

Nondimeno Dirac, spronato dal suo desiderio di semplicità, preferì inizialmente il protone per il ruolo di buca. Nel 1930 il protone e l'elettrone erano le sole particelle fondamentali osservate sperimentalmente, ed egli non era incline all'idea di introdurre una nuova entità priva di supporto sperimentale. In più, se i protoni avessero potuto essere interpretati come stati a energia negativa lasciati vacanti da elettroni, il numero delle particelle elementari si sarebbe ridotto a una sola unità: l'elettrone. E una semplificazione di questo genere, secondo Dirac, sarebbe stata «il sogno dei filosofi». Presto, però, non fu più possibile opporsi alle obiezioni a questa sua iniziale interpretazione e nel maggio 1931 egli ripiegò, non senza riluttanza, sul secondo candidato, l'antielettrone, «un nuovo tipo di particella, sconosciuto alla fisica sperimentale, che ha la stessa massa di un elettrone e carica di segno opposto». La completa simmetria tra particelle di carica positiva e di carica negativa che veniva alla luce dalla sua teoria gli diede un'ulteriore spinta a riconoscere all'antiprotone una possibilità di esistenza teorica. Quindi Dirac dovette raddoppiare il numero delle particelle elementari significative e porre le basi per speculazioni su interi mondi costituiti da antimateria.

Egli ipotizzò anche l'esistenza di un'altra particella, il monopolo magnetico, che sarebbe dotato di una carica magnetica isolata, come accade per la carica elettrica del protone e dell'elettrone. Ancora oggi non vi sono prove sperimentali certe dell'esistenza dei monopoli magnetici (si veda l'articolo *Monopoli magnetici superpesanti* di Richard A. Carrigan, Jr., e W. Peter Trower in «Le Scienze» n. 166, giugno 1982).

Nel settembre 1932 a Dirac fu affidata la Cattedra lucasiana di matematica presso l'Università di Cambridge, un incarico che Isaac Newton aveva tenuto per 30 anni e che Dirac avrebbe conservato per 37 (la cattedra è attualmente assegnata a Stephen W. Hawking). In quello stesso mese un giovane ricercatore del California Institute of Technology, Carl D. Anderson, sottopose all'attenzione della rivista «Science» un articolo che descriveva l'apparente rivelazione, nella radiazione cosmica, di «una particella dotata di carica elettrica positiva e di massa paragonabile a quella dell'elettrone». Sebbene questa scoperta non fosse affatto stata ispirata dal modello di Dirac, la nuova particella, battezzata «positrone», fu generalmente identificata con l'antielettrone di Dirac. Quando, nel 1933 a Stoccolma, fu insignito del premio Nobel, il trentunenne Dirac tenne una conferenza sulla «teoria degli elettroni e positroni». Tre anni dopo Anderson, anch'egli trentunenne, ricevette il premio Nobel per aver promosso la particella di Dirac dal semplice dominio delle ipotesi.

Elettrodinamica quantistica (QED) è il nome dato a una teoria quantistica del campo elettromagnetico. Verso la metà degli anni trenta i tentativi di formulare una soddisfacente teoria quantistica relativistica dei campi arrivarono a un punto di stallo, e molti fisici conclusero che era necessario un drastico cambiamento di tendenza. Verso la fine degli anni venti Dirac aveva dato contributi innovativi in elettrodinamica quantistica ed era dolorosamente consapevole delle insufficienze formali della struttura teorica esistente, costruita principalmente intorno a un modello proposto da Heisenberg e Pauli nel 1929. Dirac etichettò quella teoria come illogica e «turpe». Per di più, alcuni calcoli che ne facevano uso conducevano a integrali divergenti (tendenti all'infinito) cui non si poteva attribuire alcun significato fisico. Nel 1936 Dirac elaborò una teoria alternativa in cui si eliminava la conservazione dell'energia. Sebbene questa proposta radicale sia stata rapidamente confutata in sede sperimentale, Dirac continuò a criticare la teoria di Heisenberg-Pauli e a cercarne quasi ossessivamente una migliore. Nel 1979, ricordando la sua carriera, scrisse: «In realtà ho trascorso la vita tentando soprattutto di trovare equazioni migliori per l'elettrodinamica quantistica, e fin qui senza successo, ma continuo a lavorare in questa direzione».

Un percorso logico verso un'elettrodinamica quantistica perfezionata sarebbe consistito nell'utilizzare, come rampa di lancio, una migliore teoria classica dell'elettrone. Nel 1938 Dirac perseguì questa strategia e formulò una teoria relativistica classica dell'elettrone che migliorò decisamente qualla proposta da H. A. Lorentz all'inizio del secolo. La teoria di Dirac sfociò in un'equazione di moto esatta per un elettrone (trattato come una particella delle dimensioni di un punto). Poiché la teoria evitava gli infiniti e altri termini ambigui, sembrò che fosse possibile approdare a un'elettrodinamica quantistica priva di integrali divergenti. Invece, formulare una soddisfacente versione quantomeccanica della teoria classica si dimostrò più problematico di quanto Dirac avesse previsto. Egli lottò per più di vent'anni con questo problema, ma invano.

Tra il 1947 e il 1948 fu formulata una nuova teoria elettrodinamica che risolveva, in senso pratico, il problema della divergenza che in precedenza aveva compromesso i calcoli. I pionieri di questa nuova teoria, Sin-itiro Tomonaga in Giappone e Richard P. Feynman, Julian Schwinger e Freeman Dyson negli Stati Uniti, proposero un metodo detto «rinormalizzazione», in cui le grandezze che nei calcoli teorici tendevano all'infinito venivano di fatto sostituite dai valori misurati sperimentalmente della massa e della carica dell'elettrone. Questo metodo di eliminazione della divergenza rese possibili previsioni estremamente accurate, e i molti successi empirici della teoria convinsero i fisici ad adottare la rinormalizzazione come *il* metodo per l'elettrodinamica quantistica.

Dirac, però, si oppose al metodo della rinormalizzazione, giudicandolo «complicato e sgradevole» quanto la vecchia teoria di Heisenberg e Pauli. Egli riteneva che una teoria che opera con espedienti matematici *ad hoc*, non direttamente dettati da principi fisici fondamentali, non può essere buona, per quanto bene riproduca i risultati sperimentali. Le sue obiezioni, però, furono perlopiù trascurate. Negli ultimi anni fu costretto ad ammettere che non solo era stato isolato dalla comunità scientifica, ma anche che nessuna delle sue proposte per ricostruire l'elettrodinamica quantistica aveva avuto successo.

La battaglia di Dirac per una teoria quantistica dei campi alternativa diede comunque qualche significativo risultato. Uno di questi fu la sua importante teoria classica dell'elettrone, già menzionata in precedenza. Un altro fu una nuova notazione adottata in meccanica quantistica, il formalismo «bra-ket» o «bracket» (che, in inglese, significa parentesi), che introdusse elegantemente nell'argomento la potente matematica degli spazi vettoriali o spazi di Hilbert. Questo formalismo ebbe ampia diffusione grazie alla terza edizione (del 1947) dell'influente libro di testo di Dirac *The Principles of Quantum Mechanics* ed è ancora ai giorni nostri il linguaggio matematico preferito in meccanica quantistica.

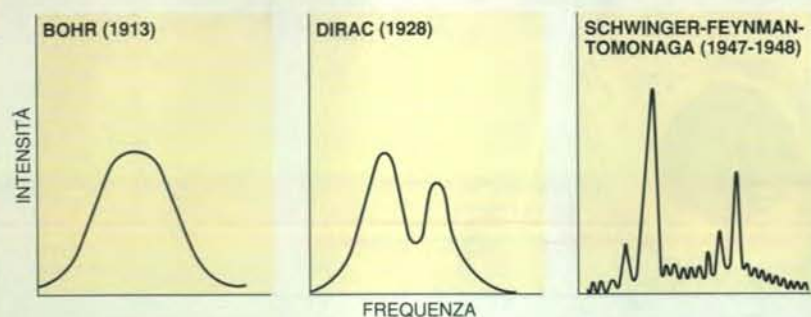
In generale, Dirac lavorò soltanto in aree piuttosto specializzate della teoria quantistica. Perciò parve in qualche modo sorprendente che, nel 1937, si avventurasse nel campo della cosmologia con una nuova idea che sviluppò in un definito modello di universo. Il suo interesse per questi argomenti era in massima parte ispirato da due dei suoi primi maestri di Cambridge, Milne ed Eddington, e dalle discussioni con un giovane talento dell'astrofisica, l'indiano Subrahmanyan Chandrasekhar, del quale fu relatore per la stesura di una parte della tesi di laurea a Cambridge.

Verso l'inizio degli anni trenta, Eddington si era imbarcato in un programma di ricerca ambizioso e non ortodosso, grazie al quale aspirava a dedurre i valori delle costanti fondamentali del-

It seems reasonable to assume that not all
the states of negative energy are occupied, but that there
are a few vacancies or 'holes' [which can be described by a wave function, like an X-ray orbit]. Such a hole would
appear experimentally as a thing with +ve energy, since
to make the hole disappear (i.e. to fill it up) one would
have to put −ve energy into it. Further, one can easily

Il concetto di antimateria, che Dirac introdusse nel 1931, si sviluppò direttamente dalla sua teoria delle «buche», qui delineata in una lettera a Niels Bohr datata 26 novembre 1929. È un esempio della chiarezza, della concisione e della grafia ordinata caratteristiche di Dirac.

Lo spettro dell'idrogeno

La riga alfa dello spettro dell'idrogeno illustra i progressi della teoria atomica da quando Niels Bohr, nel 1913, la spiegò per primo come il risultato di un singolo salto quantico. Allorché i progressi sperimentali rivelarono la struttura fine della riga, Arnold Sommerfeld combinò la teoria atomica di Bohr e la relatività speciale di Einstein per spiegare le componenti spettrali come dovute a numerose transizioni. I tentativi di ricavare i risultati di Sommerfeld dalla meccanica quantistica fallirono fino al 1928, quando la teoria dell'elettrone di Dirac riuscì a riprodurre esattamente i risultati del modello di Sommerfeld. Successive misurazioni rivelarono una struttura ancora più fine, che fu spiegata dal punto di vista teorico alla fine degli anni quaranta, grazie all'elettrodinamica quantistica moderna formulata da Julian Schwinger, Richard Feynman e Sin-itiro Tomonaga. Dirac avversò questa nuova teoria perché era, a suo parere, «solo un insieme di regole che funzionano», e non una teoria completa costruita su fondamenta «belle e solide».

la natura gettando un ponte tra la teoria quantistica e la cosmologia. Questa ricerca di una vera e propria «teoria fondamentale», come la denominò Eddington, estese l'indagine razionale al dominio della speculazione metafisica, dando luogo - come accusò un critico - a una «combinazione di paralisi della ragione e di intossicazione della fantasia». Dirac accolse con scetticismo le immaginifiche affermazioni di Eddington, ma rimase fortemente impressionato dalla sua filosofia della scienza, che sottolineava la potenza del puro ragionamento matematico, e dalla sua idea di una connessione fondamentale tra microcosmo e macrocosmo.

Nel suo primo articolo di cosmologia Dirac focalizzò l'attenzione sui grandissimi numeri «puri», ovvero adimensionali, che si possono costruire combinando le costanti fondamentali (come la costante di gravità, la costante di Planck, la velocità della luce e la massa e la carica dell'elettrone e del protone) in modo che le loro unità di misura si elidano. Egli sostenne che soltanto questi grandi numeri hanno un significato profondo in natura.

Per esempio, si sapeva che il rapporto tra la forza elettrostatica che si esercita tra un protone e un elettrone e la forza gravitazionale tra le medesime particelle è un numero molto grande, circa 10^{39}. Dirac osservò che, stranamente, questo numero era prossimo all'età allora stimata dell'universo espressa in termini di un'appropriata unità di tempo, ossia il tempo necessario alla luce per percorrere il diametro di un elettrone secondo la teoria classica.

Dirac era a conoscenza di numerose correlazioni di questo tipo tra grandi numeri puri, ma, anziché considerarle mere coincidenze, sostenne che esse costituivano l'essenza di un nuovo importante principio cosmologico, che battezzò ipotesi dei grandi numeri: «Due qualsiasi dei grandi numeri adimensionali che si osservano in natura sono connessi da una semplice relazione matematica, in cui i coefficienti sono dell'ordine di grandezza dell'unità».

A partire da questo principio Dirac giunse facilmente, ma con molte controversie, alla conclusione che la costante di gravità G è inversamente proporzionale all'età dell'universo e quindi deve essere costantemente decrescente al procedere del tempo cosmico.

Verso il 1938 Dirac aveva dedotto dall'ipotesi dei grandi numeri diverse conseguenze suscettibili di verifica sperimentale e aveva delineato il suo personale modello dell'universo basandosi su quel principio. Tuttavia la maggior parte dei fisici e degli astronomi - cui era venuto sempre più a noia l'approccio razionalista alla cosmologia - rifiutò le sue idee. Soltanto alcuni decenni più tardi, durante gli anni settanta, Dirac riprese i suoi lavori di cosmologia, perlopiù sulla base della sua teoria originaria, difendendo l'ipotesi dei grandi numeri e la previsione di una costante gravitazionale variabile dalle obiezioni basate sulle osservazioni sperimentali, e tentando di modificare il modello per adeguarlo alle nuove scoperte, come per esempio la radiazione cosmica di fondo a microonde. I suoi sforzi non gli valsero grande considerazione ed egli rimase - in cosmologia come in elettrodinamica quantistica - estraneo alle principali correnti di ricerca.

Dirac fu indissolubilmente legato al lavoro, e i colleghi lo considerarono sempre uno scapolo impenitente. Fu quindi una sorpresa quando, nel 1937, sposò Margit Wigner, sorella dell'eminente fisico ungherese Eugene Wigner. Margit era vedova, aveva avuto un figlio e una figlia dal precedente matrimonio ed ebbe altre due figlie dal matrimonio con Dirac. Come prevedibile, egli rimase avulso dalla vita familiare. «È un'ironia che solo la vita può produrre il fatto che Paul abbia sofferto a opera di suo padre, il quale aveva le stesse sue difficoltà con la famiglia» scrisse Margit. «Paul, sebbene non sia stato un padre tiranno, rimase troppo lontano dai figli. Che la storia si ripeta è semplicemente troppo vero nella famiglia Dirac.»

Dirac non coltivò mai un interesse per l'arte, la musica o la letteratura, e frequentò sporadicamente il teatro. Le sue sole passioni, fatta eccezione per la fisica, erano i viaggi e le passeggiate in montagna. Era un camminatore infaticabile e spesso nelle escursioni dimostrò doti di resistenza che meravigliarono chi lo conosceva soltanto dalle conferenze o dai banchetti ufficiali. I suoi viaggi lo portarono a fare per tre volte il giro del mondo, e a scalare alcune delle più alte vette d'Europa e d'America.

Nel settembre 1969 Dirac rinunciò alla Cattedra lucasiana e l'anno dopo decise, con la moglie, di lasciare definitivamente il Regno Unito per il clima caldo della Florida, dove ottenne un incarico presso la Florida State University a Tallahassee. Continuò a lavorare e partecipò a innumerevoli conferenze, finché la salute glielo consentì. Morì a Tallahassee nell'ottobre 1984.

BIBLIOGRAFIA

MEHRA JAGDISH e RECHENBERG HELMUT, *The Historical Development of Quantum Theory, Vol. 4, Part 1: The Fundamental Equations of Quantum Mechanics, 1925-1926*, Springer-Verlag, 1982.

DALITZ R. H. e PEIERLS SIR RUDOLF, *Paul Adrien Maurice Dirac* in «Biographical Memoirs of Fellows of the Royal Society», 32, pp. 137-185, 1986.

KURSUNOGLU BEHRAM N. e WIGNER EUGENE P. (a cura), *Reminiscences about a Great Physicist: Paul Adrien Maurice Dirac*, Cambridge University Press, 1987.

KRAGH HELGE, *Dirac: A Scientific Biography*, Cambridge University Press, 1990.